Genetic and Epigenetic Basis of Clonal Advantage during the Evolution of Acute Myeloid Leukaemia



Sven Turkalj

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

Acknowledgements

Firstly, I would like to sincerely thank my supervisor, Prof. Paresh Vyas, for his continued support of this work and scientific guidance throughout my DPhil. I am most grateful to the amazing team in the Vyas lab, who provided advice, feedback, and mentorship. In particular, Niels Asger Jakobsen has been a wonderful friend and mentor throughout my whole DPhil, who followed my first steps in the lab, taught me rigorous and precise experimental design, and who was always eager to engage in exciting scientific discussions. He has been a true role model for me. Bilyana Stoilova has also been an exceptional teacher and friend, who followed some of my first steps whilst learning computational analysis when I joined the lab. I am very grateful to her for initially setting up the ivosidenib/venetoclax project. David Cruz Hernandez has been a uniquely inspirational scientist, as we spent many evenings discussing hypotheses and fascinating biological phenomena. He also introduced me to the life in Oxford, when I first arrived in the midst of the pandemic. Angus Groom helped me a great deal with single-cell experiments; we optimized GTAC and processed numerous 384-well plates together. Felix Radtke provided exceptional bioinformatic expertise, and we had a great deal of fun outside of the lab. I am also grateful to Rabea Mecklenbrauck, who provided important clinical insights which improved my thinking about the implications of our work. A big thanks to all other past and present members of the Vyas and Porcher labs who contributed to making my DPhil a great experience: Susann Rahmig, Maria Chiara Barbanti, Gerda Mickute, David Grainger, Marlen Metzner, Batchimeg Usukhbayar, Mirian Angulo Salazar, Verena Korber, Bing Tseu, Connor Sweeney, Grigore Aristide-Gafencu, Abhimanyu Gowda, Gina Sangha, Sarah Gooding, Naser Ansari-Pour, Vasiliki Symeonidou, Rachel Moore, Vivien Ho, Arif Kirmizitas, and Emma Butterfield.

I am incredibly grateful to my co-supervisor, Prof. James O.J. Davies, for his precious scientific insights and original ideas, and for providing exceptional mentorship on the genomic aspects of my work. I would also like to thank Prof. Jim R. Hughes, for critical guidance on genomic data interpretation. Moreover, a big thank you goes to Simone G. Riva and Emine Ravza Gur from the Hughes group, for their help with bioinformatic analysis. I would sincerely like to acknowledge Prof. Ketan J. Patel and Prof. Thomas A. Milne, who have followed my scientific trajectory throughout the Transfer and Confirmation of DPhil Status; they have provided precious advice on my work, encouraging critical thinking and contributing greatly to my professional development. I would also like to

thank Prof. Claus Nerlov; by collaborating with his group, I learnt a great deal about basic haematopoietic processes.

Many of the experiments would not have been possible without amazing support from the WIMM facilities. In particular, I am so grateful to Kevin, Sally, Paul, and Craig from the Flow Cytometry Facility, who have always provided professional guidance on the complexities of FACS sorting and analysis. I am also very grateful to the staff from the Single-Cell Facility, for continued support with TARGET-seq+ and GTAC.

A huge thank you goes to friends and colleagues from the WIMM, who created a collaborative, stimulating, and exciting environment that greatly improved the quality of my work but also made the time spent in the WIMM uniquely pleasant. A special thanks goes to Matthew Baxter, Grace Meaker, Timothy Quantick, Oliver Burns, Felix Dingler, Patrik Risteski, Ramy Slama, Assunta Adamo, Affaf Aliouat, Alba Rodriguez-Meira, and Christina Simoglou Karali.

I would like to acknowledge the Scatcherd European Scholarship, the Clarendon Fund, and the Medical Research Council, for providing funding which was critical to pursue my DPhil.

Most importantly, I would like to thank my family for immense encouragement, both prior to and during my DPhil endeavours. Without the crucial support from my parents Nina and Goran, I would have never been where I am today. I am immeasurably grateful to my fiancée Arianna, for her love and patience throughout this journey. Her continued support and understanding have helped me overcome all the challenges along the way.

Finally, I would like to thank all the patients who have donated their samples for these studies. Without them, none of the work presented here would be possible.

Abstract

Acute myeloid leukaemia (AML) is an aggressive malignant disorder of immature myeloid cells in the bone marrow. The outcome of older AML patients who are unfit for intensive chemotherapy remains dismal. Relapse, which is the main cause of mortality, occurs due to Darwinian clonal selection and persistence of leukaemia stem cells (LSCs) throughout treatment. Understanding the kinetics of clonal selection and the molecular mechanisms promoting LSC phenotypes is thus urgently required.

To address the first point, I focused on *IDH1*-mutated patients enrolled in a phase Ib/II clinical trial of ivosidenib and venetoclax ± azacitidine. Despite an initial response rate of 94%, relapse was still common. To define the clonal and cellular basis of acquired resistance and sustained response, studies were conducted integrating clonal and immunophenotypic identity of single cells on longitudinal samples from 9 patients. Whilst all patients attained an initial clinical response, 7 relapsed and 2 remained in sustained remission. In all relapsing patients, therapy-resistant clones were selected early, within 3 treatment cycles, preceding relapse by months or years. Treatment failure was associated with either newly detected genetically evolved clones or impaired differentiation of pre-existing clones. In both cases, selection occurred within small LSC populations, whilst patients were in clinical remission. In patients attaining sustained response, all leukaemic clones were eradicated and rapidly replaced by clonal and wild type haematopoiesis. This study serves as proof-of-principle that early clonal selection within LSCs correlates with long-term patient outcome. If validated in larger cohorts, clonal assessment within LSCs rapidly post-treatment initiation may predict patient outcome and enable to alter treatment in those patients destined to relapse.

However, efficient therapeutic eradication of LSCs requires an understanding of how somatic mutations corrupt gene regulation to sustain LSC phenotypes. To address this gap, in the second part of this thesis, I developed a method enabling simultaneous genotyping of genomic loci with chromatin accessibility profiling at single-cell resolution, termed GTAC. In cell lines, GTAC enabled accurate and high-confidence genotyping, whilst yielding robust chromatin accessibility readouts. In human AML, GTAC efficiently mapped genetic clonal evolution to epigenomic variation, revealing chromatin changes underpinning the transformation of pre-leukaemic stem cells into *RUNX1*-mutated LSCs. Additionally, GTAC revealed that subclonal *BCOR* mutations cooperate with the *RUNX1* mutation within the leukaemic hierarchy, promoting an LSC-like epigenetic state whilst disrupting differentiation. Mechanistically, *BCOR* mutations associate with de-repression of

proximal cis-regulatory elements related to stemness genes and alter the activity of transcription factors driving self-renewal and differentiation. These analyses identified a putative therapeutic target for this patient subgroup, that warrants further functional and molecular investigation. Overall, GTAC is a robust method applicable to a broader field of cancer evolution.

Word Count: 39,450

Contents

List of Figures **xii**

List of Tables **xvi**

List of Abbreviations **xvii**

1 Introduction **1**

1.1 Normal haematopoiesis 1

1.1.1 Haematopoietic stem cells 2

1.1.2 Haematopoietic progenitor hierarchy 4

1.1.3 Transcriptional regulation of haematopoietic differentiation . 8

1.2 Malignant haematopoiesis 11

1.2.1 Acute myeloid leukaemia: an overview 11

1.2.2 Genomic landscape of AML and emerging targeted therapies 12

1.2.3 The differentiation hierarchy in AML and the leukaemia stem cell 15

1.2.4 The role of LSCs in treatment failure 20

1.2.5 The clonal hierarchy in AML 21

1.2.6 Clonal heterogeneity and AML relapse 25

1.2.7 Model systems to study the consequences of somatic mutations 28

1.2.8 Linkage of genetic and molecular landscapes in primary human samples 29

1.3 Aims 31

2 Materials and Methods **33**

2.1 Human samples and ethical approval 34

2.2 Outcome parameters 35

2.3 Cell culture 35

2.4 Bulk DNA sequencing 36

2.5 Primary sample staining for flow cytometry 37

2.6 Bulk cell sorting of primary human samples 38

2.7 TARGET-seq+ 38

2.7.1 TARGET-seq+: primer design and validation 38

2.7.2 TARGET-seq+: lysis buffer preparation, single-cell sorting, and RT-PCR 39

2.7.3 TARGET-seq+: 3' biased whole transcriptome library construction . . . 41
2.7.4 TARGET-seq+: single-cell genotyping library construction . 42
2.8 TARGET-seq+ analysis . . . 43
2.8.1 TARGET-seq+: single-cell genotyping data pre-processing . 43
2.8.2 TARGET-seq+: mutation calling . . . 43
2.8.3 TARGET-seq+: FACS index data . . . 45
2.8.4 TARGET-seq+: transcriptome data pre-processing and quality filtering . . . 45
2.8.5 TARGET-seq+: copy number alteration inference . . . 46
2.8.6 TARGET-seq+: inference of clonal hierarchies . . . 48
2.8.7 TARGET-seq+: inference of clone size in Lin^- cells . . . 49
2.9 Genotyping integration to bulk ATAC . . . 49
2.10 GTAC . . . 51
2.10.1 GTAC: primer design and validation . . . 51
2.10.2 GTAC: experimental procedure . . . 52
2.11 GTAC analysis . . . 54
2.11.1 Single-cell genotyping analysis . . . 54
2.11.2 GTAC: ADO estimation . . . 55
2.11.3 GTAC: scATAC-seq data pre-processing . . . 56
2.11.4 GTAC: quality filtering . . . 57
2.11.5 GTAC: dimensionality reduction, clustering, and cluster annotation . . . 57
2.11.6 GTAC: peak calling, marker peak identification, and co-accessibility . . . 58
2.11.7 GTAC: differential accessibility analysis . . . 59
2.11.8 GTAC: projection on reference . . . 59
2.11.9 GTAC: trajectory analysis . . . 60
2.11.10 GTAC: identification of putative PRC1.1 direct targets . . . 60
2.11.11 GTAC: TF footprinting analysis . . . 61

3 Clonal basis of resistance or response to ivosidenib combination therapies 62
3.1 Aims of Chapter 3 . . . 63
3.2 Patient cohort studied . . . 64
3.2.1 Clinical characteristics, treatment course, and responses . . . 64
3.2.2 Longitudinal immunophenotypic analysis of the patient cohort 66
3.2.3 Application of TARGET-seq+ to sequential patient samples 70
3.3 Newly detected genetic clones are selected within LSCs months or years prior to treatment failure . . . 75

3.3.1 *FLT3*-ITD or *NF1*-mutant clones emerge within LSCs 3 cycles into treatment and associate with rapid relapse 75
3.3.2 Resistant clones emerge years prior to treatment failure and expand slowly in a competing clonal milieu 79
3.4 Genetic clones detected at baseline undergo rapid therapy adaptation 82
3.4.1 A pre-exiting clone harbouring trisomy 4 loses differentiation capacity rapidly upon treatment exposure 83
3.4.2 A *FLT3*-mutated clone rapidly loses differentiation capacity upon treatment exposure 85
3.4.3 Pre-existing clones gradually lose differentiation capacity during treatment in a complex clonal milieu 87
3.5 Clonal foundations of sustained clinical response 89
3.6 Summary of Chapter 3 and proposed model 93

4 Development and validation of GTAC 97
4.1 The rationale for the development of GTAC and overview of the method 97
4.1.1 Overview of GTAC 100
4.2 Development of the GTAC chemistry 102
4.2.1 Integration of targeted genotyping to bulk ATAC 102
4.2.2 Integration of targeted genotyping to scATAC-seq in cell lines 104
4.2.3 A framework for genotyping primer testing prior to execution of GTAC 109
4.3 Validation of GTAC in cell lines 111
4.3.1 Benchmarking GTAC single-cell genotyping in cell lines . . . 111
4.3.2 Allelic drop-out estimation 115
4.3.3 Multiplexed genotyping efficiency in GTAC 117
4.3.4 Benchmarking of scATAC-seq library metrics in GTAC . . . 118
4.3.5 Integration of genotypes and chromatin accessibility profiles in cell lines 124
4.4 Brief summary of Chapter 4 125

5 GTAC links clonal evolution to epigenetic regulation in human AML 126
5.1 The main questions to address by using GTAC on human AML . . 127
5.1.1 GTAC is feasible in human BM 128
5.2 Execution of GTAC in human AML 130
5.2.1 Construction of scATAC-seq and single-cell genotyping libraries 130
5.2.2 Single-cell genotyping and scATAC-seq library metrics . . . 132

5.3 Integration of chromatin accessibility and genetic clonal identity in human AML . . . 135
5.3.1 Inference of the most likely clonal hierarchy . . . 135
5.3.2 The chromatin accessibility landscape in primary human AML 137
5.3.3 Integration of clonal identity with chromatin accessibility . . 144
5.3.4 Epigenomic correlates of *RUNX1*-R201G-associated leukaemic transformation . . . 148
5.4 Consequences of subclonal *BCOR* mutations on leukaemic differentiation . . . 155
5.4.1 Clonal basis for the LSC-like chromatin state within the leukaemic hierarchy . . . 156
5.4.2 Differentiation delay of *BCOR*-mutant leukaemic cells . . . 159
5.5 Putative epigenetic mechanisms underlying selection of *BCOR*-mutant clones . . . 164
5.5.1 Epigenetic consequences of *BCOR* mutations within the LMPP-like cluster . . . 164
5.5.2 *BCOR* mutations de-regulate promoters of LSC-associated genes . . . 170
5.5.3 Aberrant TF activity in *BCOR*-mutant LMPP-like cells . . . 173
5.5.4 Epigenetic heterogeneity between genetically similar clones . 177
5.6 Brief summary of Chapter 5 . . . 179

6 Discussion **182**
6.1 The implications of clonal heterogeneity for AML treatment . . . 182
6.2 Clonal and cellular basis of resistance and response to ivosidenib combination therapies . . . 185
6.2.1 Rapid clonal selection within LSCs foreshadows patient outcome 185
6.2.2 Heterogeneous genetic and cellular routes towards relapse . . 187
6.2.3 Variable duration of clinical response despite universally rapid clonal selection . . . 188
6.2.4 Implications for clinical practice . . . 189
6.2.5 Challenges ahead and future outlook . . . 190
6.3 GTAC: integrating clonal identity with chromatin accessibility . . . 191
6.3.1 Technical improvement of GTAC over prior approaches . . . 191
6.3.2 Insights into *RUNX1*-associated leukaemic transformation . 192
6.3.3 The impact of somatic mutations on LSC gene regulation . . 194
6.3.4 Future applications of GTAC . . . 196
6.3.5 Technical considerations, limitations, and future outlook . . 198

Appendices

A Appendices **200**
A.1 An Overview of Targeted Therapies in Acute Myeloid Leukemia . . 201
A.2 Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging . 212
A.3 GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells 266
A.4 A protocol for simultaneous high-sensitivity genotyping and chromatin accessibility profiling in single cells 312
A.5 Mutations identified by bulk targeted panels and WES at baseline (related to Chapter 3) 366
A.6 Mutations identified by bulk targeted panels and WES at relapse (related to Chapter 3) 367
A.7 Pre-amplification primers used for TARGET-seq+ (related to Chapter 3) 368
A.8 Nested (PCR1) primers used for TARGET-seq+ (related to Chapter 3) 369
A.9 Rates of successful clonal assignment per sample (related to Chapter 3) 370
A.10 inferCNV genome-wide for pt15 (related to Chapter 3) 371
A.11 inferCNV genome-wide for pt5 (related to Chapter 3) 372
A.12 inferCNV genome-wide for pt14 (related to Chapter 3) 373
A.13 inferCNV genome-wide for pt9 (related to Chapter 3) 374
A.14 inferCNV genome-wide for pt4 (related to Chapter 3) 375
A.15 Inferred clone sizes within all live, Lin$^-$ cells for pt11 (related to Chapter 3) 376
A.16 Inferred clone sizes within all live, Lin$^-$ cells for pt5 (related to Chapter 3) 377
A.17 Inferred clone sizes within all live, Lin$^-$ cells for pt14 (related to Chapter 3) 378
A.18 Inferred clone sizes within all live, Lin$^-$ cells for pt9 (related to Chapter 3) 379
A.19 Inferred clone sizes within all live, Lin$^-$ cells for pt4 (related to Chapter 3) 380
A.20 Inferred clone sizes within all live, Lin$^-$ cells for pt18 (related to Chapter 3) 381
A.21 Inferred clone sizes within all live, Lin$^-$ cells for pt10 (related to Chapter 3) 382
A.22 Single-cell genotyping for pt11 (related to Chapter 3) 383
A.23 Single-cell genotyping for pt15 (related to Chapter 3) 384
A.24 Single-cell genotyping for pt20 (related to Chapter 3) 385
A.25 Single-cell genotyping for pt5 (related to Chapter 3) 386

A.26 Single-cell genotyping for pt14 (related to Chapter 3) 387
A.27 Single-cell genotyping for pt9 (related to Chapter 3) 388
A.28 Single-cell genotyping for pt4 (related to Chapter 3) 389
A.29 Single-cell genotyping for pt18 (related to Chapter 3) 390
A.30 Single-cell genotyping for pt10 (related to Chapter 3) 391
A.31 Genotyping primers used for GTAC (related to Chapters 4 and 5) . 392

References **393**

List of Figures

1.1 Haematopoietic trajectory 5
1.2 Continuum of lympho-myeloid progenitors 7
1.3 Targeted therapies in AML. 13
1.4 LMPP-like and GMP-like LSC hierarchy. 18
1.5 Link between clonal and differentiation hierarchies in AML. 22

3.1 Patient cohort studied and treatment responses. 65
3.2 Representative FACS panels. 67
3.3 FACS analysis of HSPC, GM precursor, and mature myeloid compartments. 69
3.4 FACS analysis of HSPC sub-compartments. 70
3.5 TARGET-seq+ framework for clonal, transcriptomic, and immunophenotypic single-cell analysis of sequential patient samples. 71
3.6 Single-cell genotyping primer testing for TARGET-seq+. 73
3.6 Single-cell genotyping primer testing for TARGET-seq+. 74
3.7 TARGET-seq+ libraries generated from a patient sample. 74
3.8 Longitudinal clonal analysis for pt11. 76
3.9 Longitudinal clonal analysis for pt15. 78
3.10 Longitudinal clonal analysis for pt20. 80
3.11 Longitudinal clonal analysis for pt5. 82
3.12 Longitudinal clonal analysis for pt14. 84
3.13 Longitudinal clonal analysis for pt9. 86
3.14 Longitudinal clonal analysis for pt4. 88
3.15 Longitudinal clonal analysis for pt18. 90
3.16 Longitudinal clonal analysis for pt10. 92
3.17 Rapid clonal selection within specific immunophenotypic cell populations. 94
3.18 Correlation of resistance-associated somatic aberration and response to treatment. 95
3.19 Rapid clonal selection within immature BM cells foreshadows patient outcome. 96

4.1 Advantages of understanding clone-specific chromatin accessibility. . 99
4.2 Schematic representation of the GTAC method. 101
4.3 Integration of targeted genotyping to bulk ATAC. 103

4.4 PCR cycle number determination in scATAC-seq. 105
4.5 Optimization of genotyping primer concentration. 106
4.6 GTAC scATAC library traces at single-cell resolution. 107
4.7 Multiplexed single-cell genotyping testing in GTAC. 108
4.8 Framework for single-cell genotyping primer testing and optimization in GTAC. 110
4.9 Genotyping locus capture in GTAC. 112
4.10 Mutation zygosity estimation in GTAC single-cell genotyping. . . . 114
4.11 ADO estimation in GTAC. 116
4.11 ADO estimation in GTAC. 117
4.12 Multiplexed genotyping success rates in K562. 118
4.13 GTAC with increasing numbers of genotyped loci. 119
4.14 scATAC-seq library metrics in GTAC. 120
4.14 scATAC-seq library metrics in GTAC. 121
4.15 Correlation of scATAC-seq libraries. 122
4.16 scATAC-seq coverage at genotyped loci. 123
4.17 K562 cells passing quality filters in GTAC. 124
4.18 Cell line mixing experiment. 125

5.1 GTAC testing in primary human BM. 129
5.2 GTAC on primary human AML. 131
5.3 ADO estimation in the AML sample. 133
5.4 scATAC-seq metrics in the AML sample. 134
5.5 AML cells passing GTAC quality filters. 135
5.6 Single-cell genotyping of the AML sample. 136
5.7 Clonal hierarchy reconstruction. 138
5.8 Chromatin accessibility landscape of the AML sample. 139
5.9 scATAC-seq cluster annotation. 140
5.9 scATAC-seq cluster annotation. 141
5.10 Mapping to a scATAC-seq reference and cluster-specific chromatin accessibility. 143
5.10 Mapping to a scATAC-seq reference and cluster-specific chromatin accessibility. 144
5.11 Integration of genotypes with chromatin accessibility. 145
5.11 Integration of genotypes with chromatin accessibility. 146
5.12 Clone-specific TF binding motif accessibility. 147
5.13 Chromatin differences between pre-leukaemic and leukaemic genetic clones. 149
5.13 Chromatin differences between pre-leukaemic and leukaemic genetic clones. 150

5.14 Differentiation stage at which leukaemic transformation occurred. . 151
5.15 Two proposed models for the COO of the leukaemic clones 152
5.16 Selected chromatin differences between pre-leukaemic and leukaemic LMPPs. 153
5.17 Leukaemic LMPPs display a pro-proliferative chromatin signature. . 154
5.18 Epigenetic resemblance of the LMPP-like cluster to LSCs. 157
5.18 Epigenetic resemblance of the LMPP-like cluster to LSCs. 158
5.19 Differentiation trajectories within the leukaemic hierarchy. 159
5.20 Clonal frequency across the leukaemic hierarchy. 160
5.21 Clone-specific TF binding motif accessibility through pseudotime. . 161
5.22 Unbiased analysis of TF motif accessibility through pseudotime. . . 162
5.23 Clone-specific chromatin changes associated with myeloid differentiation. 163
5.24 Differentially accessible genes between *BCOR*-mutant and *BCOR*-WT LMPP-like cells. 165
5.24 Differentially accessible genes between *BCOR*-mutant and *BCOR*-WT LMPP-like cells. 166
5.25 Chromatin characteristics of *BCOR*-mutant LMPP-like cells. 167
5.26 Clone-specific *INPP4B* accessibility. 168
5.27 Clone-specific gene accessibility through pseudotime. 169
5.27 Clone-specific gene accessibility through pseudotime. 170
5.28 *INPP4B* promoter de-regulation in *BCOR*-mutant LMPP-like cells. 171
5.29 Promoter de-regulation in *BCOR*-mutant LMPP-like cells. 172
5.30 TF binding motif accessibility between *BCOR*-mutant and *BCOR*-WT LMPP-like cells. 174
5.30 TF binding motif accessibility between *BCOR*-mutant and *BCOR*-WT LMPP-like cells. 175
5.31 TF binding motif accessibility between *BCOR*-mutant and *BCOR*-WT Mono/Neut progenitor-like cells. 176
5.32 Simulation of ADO effect on differential accessibility analysis. . . . 177
5.33 Differential gene accessibility between *BCOR*-mutant clones. 178
5.34 Differential TF binding motif accessibility between the two *BCOR*-mutant clones. 179
5.35 Clonal evolution correlates with epigenetic change. 181

A.1 inferCNV genome-wide for pt15 (related to Chapter 3). 371
A.2 inferCNV genome-wide for pt5 (related to Chapter 3). 372
A.3 inferCNV genome-wide for pt14 (related to Chapter 3). 373
A.4 inferCNV genome-wide for pt9 (related to Chapter 3). 374
A.5 inferCNV genome-wide for pt4 (related to Chapter 3). 375

A.6 Single-cell genotyping for pt11. 383
A.7 Single-cell genotyping for pt15. 384
A.8 Single-cell genotyping for pt20. 385
A.9 Single-cell genotyping for pt5. 386
A.10 Single-cell genotyping for pt14. 387
A.11 Single-cell genotyping for pt9. 388
A.12 Single-cell genotyping for pt4. 389
A.13 Single-cell genotyping for pt18. 390
A.14 Single-cell genotyping for pt10. 391

List of Tables

3.1 Patient characteristics. 65
3.2 FACS analysis of normal BM. 68
3.3 Longitudinal karyotyping. 72

4.1 Genomic loci screened by GTAC in cell lines. 111

5.1 Somatic mutations and germline heterozygous SNPs genotyped by GTAC in the primary human AML sample. 130

A.1 Mutations identified by bulk targeted panels and WES at baseline (related to Chapter 3). 366
A.2 Mutations identified by bulk targeted panels and WES at relapse (related to Chapter 3). 367
A.3 Pre-amplification primers used for TARGET-seq+ (related to Chapter 3). 368
A.4 Nested (PCR1) primers used for TARGET-seq+ (related to Chapter 3). 369
A.5 Rates of successful clonal assignment per sample (related to Chapter 3). 370
A.6 Inferred clone sizes within all live, Lin$^-$ cells for pt11 (related to Chapter 3). 376
A.7 Inferred clone sizes within all live, Lin$^-$ cells for pt5 (related to Chapter 3). 377
A.8 Inferred clone sizes within all live, Lin$^-$ cells for pt14 (related to Chapter 3). 378
A.9 Inferred clone sizes within all live, Lin$^-$ cells for pt9 (related to Chapter 3). 379
A.10 Inferred clone sizes within all live, Lin$^-$ cells for pt4 (related to Chapter 3). 380
A.11 Inferred clone sizes within all live, Lin$^-$ cells for pt18 (related to Chapter 3). 381
A.12 Inferred clone sizes within all live, Lin$^-$ cells for pt10 (related to Chapter 3). 382
A.13 Genotyping primers used for GTAC (related to Chapters 4 and 5). . 392

List of Abbreviations

2-HG 2-hydroxyglutarate

ADO . . . . . Allelic drop-out

∝-KG . . . . . ∝-ketoglutarate

ALAL . . . . . Acute leukaemia of ambiguous lineage

AML . . . . . Acute myeloid leukaemia

AML-CFU . . AML colony forming unit

APL . . . . . Acute promyelocytic leukaemia

ATAC-seq . . Assay for transposase-accessible chromatin with sequencing

bp . . . . . Base pair

BM . . . . . Bone marrow

BNKp . . . . . B-cell/natural killer cell progenitor

BSA . . . . . Bovine serum albumine

cDNA . . . . . Complementary DNA

CH . . . . . Clonal haematopoiesis

ChIP-seq . . . Chromatin immunoprecipitation with sequencing

CI . . . . . Confidence interval

CLP . . . . . Common lymphoid progenitor

CML . . . . . Chronic myelogeneous leukaemia

CMP . . . . . Common myeloid progenitor

CNA . . . . . Copy number alteration

COO . . . . . Cell of origin

COSMIC . . . Catalogue of Somatic Mutations in Cancer

CR . . . . . Complete remission

CRi . . . . . CR with incomplete haematologic recovery

CSC . . . . . Cancer stem cell

DC . . . . . Dendritic cell

DNA . . . . . Deoxyribonucleic acid

DOR . . . . . Duration of response

DTA . . . . . *DNMT3A*, *TET2*, and *ASXL1*

DTT 1,4-dithiothreitol

ECOG . . . . . Eastern Cooperative Oncology Group

EDTA . . . . . Ethylenediaminetetraacetic acid

ELN . . . . . European LeukemiaNet

EMPP . . . . . Erythroid/megakaryocyte-primed multipotent progenitor

EOC . . . . . End of cycle

FACS . . . . . Fluorescence-activated cell sorting

FBS . . . . . Fetal bovine serum

FRiP . . . . . Fraction of reads in peaks

FC . . . . . Fold change

FDR . . . . . False discovery rate

FMO . . . . . Fluorescence-minus-one

gDNA . . . . . Genomic DNA

GM . . . . . granulocyte-macrophage

GMP . . . . . Granulocyte-monocyte progenitor

GoF . . . . . Gain-of-function

GoT . . . . . Genotyping of transcriptomes

GoTChA . . . Genotyping of targeted loci with single-cell chromatin accessibility

GSEA . . . . . Gene set enrichment analysis

GTAC . . . . . Genotyping with the assay for transposase-accessible chromatin

GTP . . . . . Guanosine triphosphate

H2AK119Ub1 Monoubiquitin mark on Lysine 119 of the histone H2A

hg . . . . . Human genome build

HSC . . . . . Haematopoietic stem cell

HSPC . . . . . Haematopoietic stem and progenitor cell

IDH . . . . . Isocitrate dehydrogenase

IMDM . . . . Iscove's Modified Dulbecco's Medium

Indel . . . . . Small insertion or deletion

Lin . . . . . Lineage

LMPP . . . . . Lymphoid-primed multipotent progenitor

LoF . . . . . Loss-of-function

LSC . . . . . Leukaemia stem cell

LSI . . . . . Latent semantic indexing

MARCH . . . Mechanisms of age-related clonal haematopoiesis

MDS . . . . . Myelodysplastic syndrome

MDS/MPN . Myelodysplastic syndrome/myeloproliferative neoplasm overlap syndrome

MEP . . . . . Megakaryocyte-erythroid progenitor

MkP . . . . . Megakaryocyte progenitor

***MLL*r** . . . . . *MLL*-rearranged

MLP . . . . . Multi-lymphoid progenitor

MNC . . . . . Mononuclear cell

MPAL . . . . . Mixed-phenotype acute leukaemia

MPN . . . . . Myeloproliferative neoplasm

MPP . . . . . Multipotent progenitor

MRD . . . . . Measurable residual disease

mRNA . . . . Messenger ribonucleic acid (RNA)

NGS . . . . . Next-generation sequencing

NK . . . . . Natural killer

NKP . . . . . Natural killer cell progenitor

NOC . . . . . Nuffield Orthopaedic Centre

ORR . . . . . Overall response rate

OS . . . . . Overall survival

PB . . . . . Peripheral blood

PBMC . . . . Peripheral blood mononuclear cell

PBS . . . . . Phosphate-buffered saline

PCR . . . . . Polymerase chain reaction

pDC . . . . . Plasmacytoid dendritic cell

polyA . . . . . Polyadenylated

PR . . . . . Partial response

PRC . . . . . Polycomb repressive complex

pt . . . . . Patient

QC . . . . . Quality control

qPCR . . . . . Quantitative PCR

r . . . . . Pearson correlation coefficient

R/R . . . . . Relapsed/refractory

RNA . . . . . Ribonucleic acid

RNA-seq . . . RNA sequencing

RPMI-1640 . Roswell Park Memorial Institute medium

RT . . . . . Reverse transcription

RT-PCR . . . Reverse transcription followed by PCR

RTK . . . . . Receptor tyrosine kinase

sAML . . . . . Secondary acute myeloid leukaemia

scATAC-seq . Single-cell ATAC-seq

SCID . . . . . Severe combined immunodeficient

scRNA-Seq . . Single-cell RNA-seq

scVAF . . . . . Single-cell variant allele frequency

SD . . . . . Standard deviation

SL-IC . . . . . SCID leukaemia-initiating cell

SNP . . . . . Single nucleotide polymorphism

SNV . . . . . Single nucleotide variant

TF . . . . . Transcription factor

TS+ . . . . . TARGET-seq+

TSS . . . . . Transcription starting site

TSO . . . . . Template-switching oligo

UMAP . . . . Uniform manifold approximation and projection

VAF . . . . . Variant allele frequency

WES . . . . . Whole-exome sequencing

WT . . . . . Wild type

Chapter 1

Introduction

Contents

1.1 Normal haematopoiesis 1
1.1.1 Haematopoietic stem cells 2
1.1.2 Haematopoietic progenitor hierarchy 4
1.1.3 Transcriptional regulation of haematopoietic differentiation 8
1.2 Malignant haematopoiesis 11
1.2.1 Acute myeloid leukaemia: an overview 11
1.2.2 Genomic landscape of AML and emerging targeted therapies 12
1.2.3 The differentiation hierarchy in AML and the leukaemia stem cell 15
1.2.4 The role of LSCs in treatment failure 20
1.2.5 The clonal hierarchy in AML 21
1.2.6 Clonal heterogeneity and AML relapse 25
1.2.7 Model systems to study the consequences of somatic mutations 28
1.2.8 Linkage of genetic and molecular landscapes in primary human samples 29
1.3 Aims 31

1.1 Normal haematopoiesis

Every day, the adult human bone marrow (BM) generates $\approx 10^{11}$ new mature blood cells [1–3] through the process of haematopoietic differentiation (numbers estimated via the total blood cell counts and lifespans of different cell types). Life-long haematopoiesis is sustained by a rare population of self-renewing haematopoietic stem cells (HSC), located at the apex of haematopoietic differentiation. Downstream of HSCs, a hierarchy of increasingly committed progenitors gives rise to all blood lineages. These progenitors act as transit-amplifying cells, which gain proliferative potential whilst gradually losing self-renewal capacity, as they differentiate towards terminally mature states. Decades of functional and molecular investigation have revolutionized our understanding of haematopoiesis. However, important

questions persist, regarding the mechanisms regulating HSC maintenance, lineage fate decisions, and hierarchical relations between progenitor populations. Here, our current understanding of normal haematopoiesis is reviewed first, followed by a description of how this process gets altered in myeloid malignancy.

1.1.1 Haematopoietic stem cells

An adult human BM is estimated to contain between 50,000 and 200,000 HSCs [4]. The HSC is a rare, undifferentiated cell, defined by three features. Firstly, it possesses self-renewing potential [5, 6], a unique capability of stem cells to generate daughter stem cells, thus maintaining the stem cell pool. Secondly, HSCs reside in a dormant, quiescent state [7], which limits active entry into the cell cycle, genomic damage, and exhaustion of the HSC pool. Thirdly, the HSC is multipotent, meaning it can give rise to all blood lineages [5, 6]. The initial evidence for the existence of HSCs came in the early 1960s, when unfractionated BM was transplanted into lethally irradiated mice [8]. The generation of clonal splenic colonies demonstrated that cells from different lineages can arise from a single multipotent cell. In 1988, using flow-activated cell sorting (FACS), the Weissman group identified that a Thy-1^{lo}Sca-1^{+} BM fraction depleted for maturation markers could reconstitute the BM of lethally irradiated recipients and give rise to T-cell, B-cell, and myeloid lineages, thus identifying murine HSCs [9]. In the same year, the Dick group established a xenotransplantation model, in which human BM could be transplanted into immunodeficient mice [10], paving the way for the *in vivo* study of human HSCs. Subsequently, both murine and human multipotent cells capable of long-term multi-lineage engraftment have been purified [5, 6, 11]. Human HSCs are immunophenotypically defined as Lineage^{-}(Lin^{-})CD34^{+}CD38^{-}CD90^{+} cells [5, 6], a classification which was further refined by using CD45RA and CD49f surface markers [12]. Accordingly, human long-term HSCs reside in the Lin^{-}CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}CD49f^{+} fraction. Prospective purification of HSCs enabled their detailed functional characterization and identified the major regulators of the HSC state.

At any time during unperturbed haematopoiesis, most HSCs exist in the G_0 stage of the cell cycle, thus residing in a dormant state [7, 13]. Dormancy is characterized by low levels of the transcription factor (TF) c-Myc [14], absent CDK6 activity [15], and depends on negative cell cycle regulators like p57 and p18 [16, 17]. Quiescent HSCs display a low metabolic profile, characterized by low rates of protein synthesis [18], whilst relying on autophagy [19], mitophagy [20], and a high lysosomal content [21–23]. This HSC-specific metabolic milieu protects the HSC pool from endogenous damage and exhaustion.

The quiescent state is intertwined with the ability of HSCs to self-renew. These processes are regulated both extrinsically, by signals from the HSC niche, and intrinsically, by epigenetic regulators and TFs. Although the precise molecular circuitries are far from fully understood, xenotransplantation assays coupled to genetic perturbation identified multiple genes critical for the maintenance of HSC self-renewal and quiescence. These include TFs like ETV6 [24], ERG [25], FLI-1 [26], PU.1 (SPI-1) [27], MECOM [28–30], GATA-2 [31, 32], HLF [33], GFI-1 [34, 35], NR4A2 [36], PBX1 [37], c-MYB [38], STAT5 [39], RELA/B [40, 41], NF-Y [42, 43], and FOXO and HOX TF families [44–48]. Furthermore, MPL [49] and KIT [50] cytokine receptors, WNT signalling through β-catenin [51], and a component of the superelongation complex, MLLT3 [52], are critical regulators of HSC self-renewal.

The HSC-specific chromatin landscape governs HSC self-renewal, and several epigenetic regulators influence the decisions between HSC self-renewal and differentiation. The histone methyltransferase MLL1, crucial for HSC maintenance [53, 54], promotes the expression of *HOX* genes [53, 55] and other HSC regulators. Polycomb repressive complexes PRC1 and PRC2, which antagonize gene expression via deposition of repressive histone marks, are also extensively implicated in HSC regulation. Genetic deletion of *Bmi-1* and *Eed*, components of PRC1 and PRC2, respectively, and overexpression of *Ezh2*, encoding the catalytic component of PRC2, demonstrated that PRC1/2 are required for HSC self-renewal [56–59]. However, the positive regulation of HSC self-renewal by PRC2 has been challenged by other studies [60, 61], which instead suggested PRC2 limits HSC self-renewal.

The decisions between HSC maintenance and differentiation are also regulated by DNA methylation. DNMT3A and TET2, which catalyse DNA methylation and hydroxymethylation, respectively, exhibit a limiting effect on HSC properties, as their deletion enhances self-renewal [62–66]. Besides proximal gene regulation, the quiescent HSC state is also wired at the level of 3D chromatin structure, as CTCF-induced genome looping accompanies HSC activation [67].

The HSC population is heterogeneous in terms of dormancy levels, lineage output, and response to external stimuli. HSCs exist in a continuum of cell cycle states, ranging from highly dormant to cycle-primed [17]. Moreover, although HSCs are multipotent at the population level, single HSCs are differentially primed towards different lineages. Single murine HSCs differ in respect to their lymphoid and myeloid output [68, 69], and HSCs with a bias for platelet production were identified in mice and humans [70–75]. Recent molecular analyses suggest the differential fate priming within HSCs is encoded both transcriptionally and epigenetically [75, 76]. Interestingly, both the lineage bias and the overall contribution to blood production by HSCs vary with age [69, 75, 77, 78]. Finally, distinct HSC sub-populations displaying variable response to inflammation have recently been identified in mice and humans [79, 80]. Hence, the heterogeneous HSC pool is likely regulated by a combination of intrinsic factors and external pressures.

1.1.2 Haematopoietic progenitor hierarchy

Downstream of the HSC, a continuum of highly proliferative progenitors undergoes gradual commitment towards mature blood lineages. Historically, this process was believed to proceed through discrete progenitor stages, where stepwise lineage decisions were accompanied by increase in commitment and decrease in cellular potency. However, this model has been heavily debated, with the current understanding being that unilineage-restricted progenitors stem from a continuum of undifferentiated (but primed) haematopoietic stem and progenitor cells (HSPCs) [81–83](Figure 1.1). Hence, whilst distinct progenitor populations can generate multiple lineages, most single cells within these populations are destined towards a single fate, with

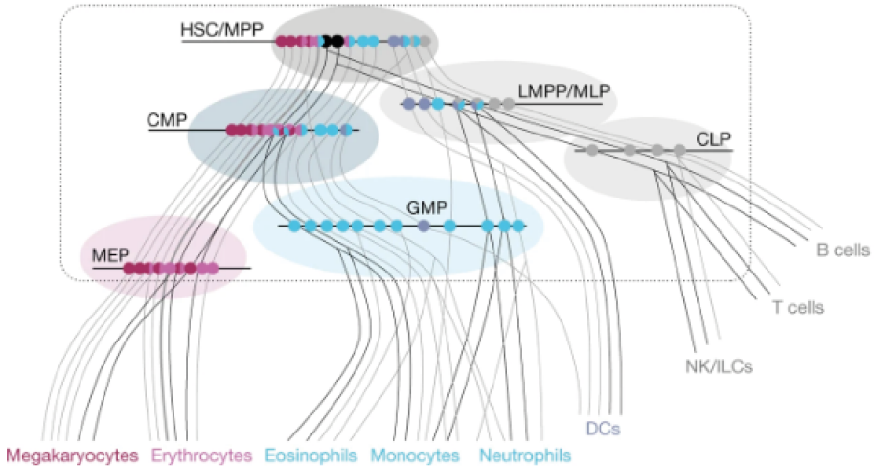


Figure 1.1: Haematopoietic trajectory. Colours indicate different lineages. Shaded areas indicate cell populations commonly defined by surface markers. Reprinted from [84], with permission from Springer Nature.

the lineage-related cellular identity gradually increasing through differentiation. The haematopoietic cascade initiates as HSCs give rise to multi-potent progenitors (MPP). In 2007, using human cord blood, the Weissman group defined the MPP as the $Lin^{-}CD34^{+}CD38^{-}CD45RA^{-}CD90^{-}$ cell fraction, which arises from HSCs both *in vitro* and *in vivo* [85]. Whilst still capable of generating multi-lineage engraftment, MPPs have lower primary and secondary engraftment capacity compared to HSCs [85–87]. MPPs resemble HSCs in terms of transcriptome [80, 88–91], proteome [87], and chromatin accessibility [88, 92], suggesting they inherit most HSC programs. In particular, MPPs resemble murine activated short-term HSCs [80, 93, 94] and, accordingly, upregulate the cell cycle machinery [87]. Although multipotent as population, MPP sub-populations are lineage-biased [70, 86, 95, 96]. Indeed, it was proposed that the first lineage decisions occur within MPPs [97], although we now know that this can occur as early as the HSC (reviewed above).

Downstream of the MPP, the first major segregation in lineage potential at population level occurs between the lymphoid-primed multipotent progenitors (LMPP), restricted to lymphoid, neutrophil, monocyte, and dendritic cell (DC) fates [80, 90, 98–100], and the erythroid/megakaryocyte-primed multipotent progenitor (EMPP), restricted to erythroid, megakaryocytic, mast cell, basophil, and eosinophil

fates [80, 89]. This contrasts the classic view that the first lineage branching occurs between the lymphoid and myeloid lineages, through the common myeloid and lymphoid progenitors (CMP and CLP, respectively). Instead, the LMPP/EMPP model is consistent with early segregation between the myelo-erythroid (EMPP) and myelo-lymphoid (LMPP) fates [101, 102]. The EMPP subsequently gives rise to megakaryocyte-erythroid progenitors (MEP), eosinophil progenitors and basophil-mast cell progenitors [89, 103]. Given the relevance of the LMPP in the development of acute myeloid leukaemia (AML), the remainder of this section will instead focus on the LMPP and its progeny.

The LMPP was first identified in mice by the Jacobsen group [98]. This work showed that $Flt3^+$ murine immunophenotypic 'HSCs' generate lymphoid cells, granulocytes, and monocytes, but not erythroid and megakaryocytic cells, *in vitro* and *in vivo*. Consistently, the transition from HSC to LMPPs is accompanied by downregulation of megakaryocytic/erythroid genes, with concurrent upregulation of lymphoid/granulocytic/monocytic genes [99]. Further work showed that murine LMPPs generated DCs and that most single LMPPs harboured uni-lineage commitment [104]. In humans, LMPPs are immunophenotypically defined as $Lin^-CD34^+CD38^-CD10^-CD90^-CD45RA^+$ [100] and arise from MPPs [85]. The first comprehensive characterization of human LMPPs was performed in 2011 by the Vyas group [100]. Here, colony assays demonstrated that LMPPs had lympho/myeloid, but not erythroid/megakaryocytic potential, and demonstrated replating potential intermediate between MPPs and downstream progenitors. Consistent with murine studies, profiling of a selected subset of genes suggested lympho/myeloid as opposed to erythroid/megakaryocytic priming within human LMPPs. Further evidence for human LMPPs suggested this $CD10^-$ population was upstream of $CD10^+$ lymphoid progenitors and could give rise to monocytic but not erythroid output [105]. This was corroborated by molecular profiling [90], which revealed human LMPPs form a transcriptional continuum with two more committed progenitor populations, namely $Lin^-CD34^+CD38^+CD10^-CD123^+CD45RA^+$ granulocyte-macrophage progenitors (GMP), and $Lin^-CD34^+CD38^-CD10^+CD90^{-/lo}CD45RA^+$ multi-lymphoid

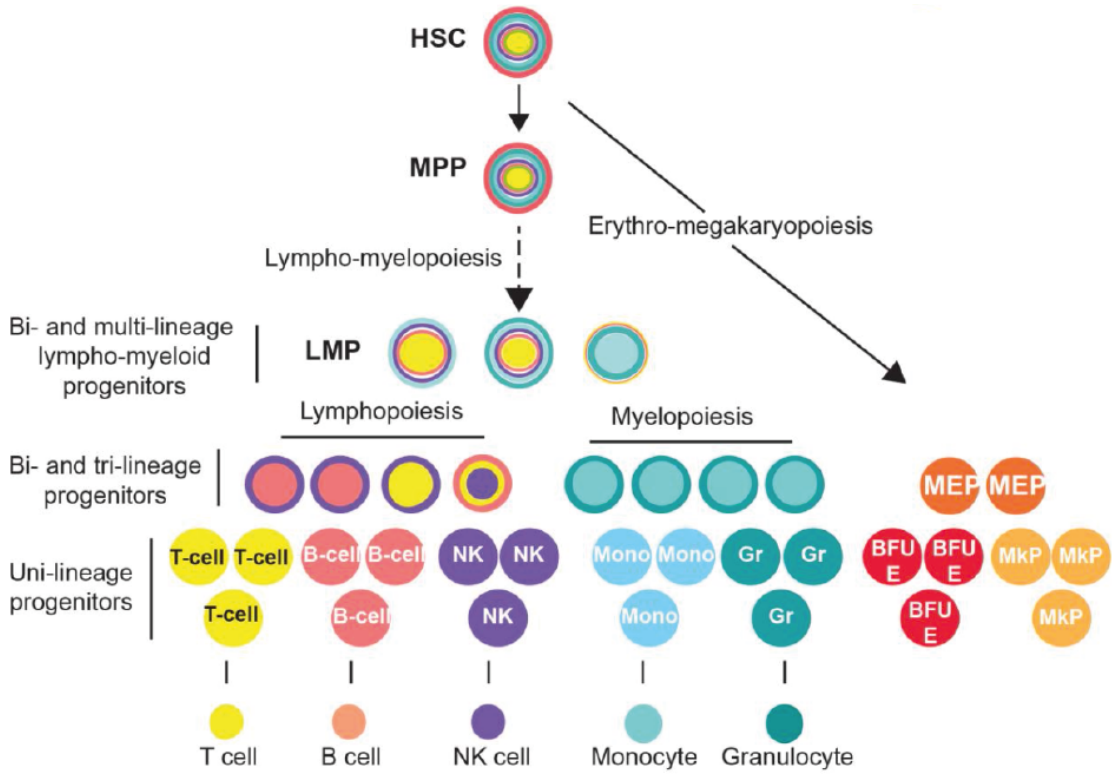


Figure 1.2: Continuum of lympho-myeloid progenitors. Here, LMP, and not LMPP, is used to highlight that not all LMPPs are lymphoid-primed. LMPPs form a continuum with progenitors harbouring predominantly lymphoid (MLP; yellow/pink/purple) or myeloid (GMP; blue/cyan) potential. Reprinted from [90], with permission from Springer Nature.

progenitors (MLP) (Figure 1.2). Functionally, GMPs give rise to neutrophils and monocytes/macrophages, with residual lymphoid output [90], whilst MLPs mostly produce lymphoid cells, with residual monocytic/macrophage/DC, but not granulocytic output [90, 106]. MLPs further differentiate into $Lin^{-}CD34^{+}CD38^{+}CD10^{+}CD90^{-}CD45RA^{+}$ B/NK cell progenitors (BNKp), devoid of myeloid potential [90, 107]. The predominantly myeloid GMP [90, 100, 108, 109] differentiates into promyelocyte-like $CD34^{-}CD244^{+}CD117^{+}$ granulocyte-macrophage precursors (GM precursors), which generate $CD34^{-}CD244^{+}CD117^{-}$ granulocytes and monocytes [110]. Compared to $CD34^{+}$ progenitors, $CD34^{-}$ GM precursors give rise only to short-lived colonies *in vitro*, consistent with loss of progenitor potential.

In most AML patients, LMPPs, GMPs, and/or GM precursors are abnormally expanded, indicating high transforming potential within these compartments. Inter-

estingly, immunophenotypic LMPPs are extremely rare in normal BM, constituting only around 0.02% of mononuclear cells (MNCs) [90]. Conversely, in AML, LMPPs can comprise up to 80% of BM MNCs [100]. Hence, the physiologic size of the LMPP may be rigorously controlled to minimize the risk of uncontrolled malignant proliferation.

1.1.3 Transcriptional regulation of haematopoietic differentiation

Haematopoietic differentiation is regulated by the coordinated activity of lineage-affiliated TFs. In conjunction with epigenetic regulators, TFs translate extrinsic and intrinsic signals into chromatin changes which promote commitment towards one lineage whilst restricting alternative fates. TFs which are active within the earliest HSPCs coordinate lineage priming, thus promoting a state which is permissive towards one or few lineages. Within downstream progenitors, additional TFs then drive lineage specification and commitment, which ultimately lock the differentiation pathway towards one specific fate (reviewed in [111, 112]). Importantly, genes encoding these TFs are subject to recurrent genetic alterations in blood cancer, highlighting their critical role for normal haematopoietic differentiation. Decades of research focused on lineage-affiliated TFs have been extensively reviewed in [111–115]. Here, I provide a brief overview of some of the best characterized lineage-affiliated TFs, focusing primarily on lympho-myeloid differentiation.

Besides its role in HSC maintenance [27], PU.1, a member of the ETS-family of TFs, is a master regulator of lympho-myeloid lineage priming and differentiation [116, 117]. PU.1 binds to regulatory sequences of myeloid and lymphoid genes, including several cytokine receptors [118], allowing early progenitors to sense cytokine signals promoting lymphoid/myeloid fates. Deletion of PU.1 in adult BM causes a loss of myeloid and lymphoid progenitors [27, 119] and PU.1-deficient GMPs cannot complete myeloid maturation [27]. Hence, PU.1 drives lineage priming of MPPs and LMPPs, and differentiation from oligopotent GMPs. PU.1, like other lineage-affiliated TFs, not only stimulates genes from a specific lineage, but concurrently antagonizes alternative fates. When PU.1 is overexpressed in

multipotent progenitors, this promotes downregulation of GATA-1 [120], a TF which is the master regulator of mature red blood cell [121, 122], megakaryocyte [123, 124], eosinophil [125], basophil [126], and mast cell [127] lineages. Accordingly, GATA-1 overexpression in HSCs abolishes lymphoid/myeloid fates [128]. MPPs expressing high levels of PU.1 have granulocyte/monocyte/lymphoid potential, in contrast to GATA-1 expressing MPPs [102]. Accordingly, PU.1 and GATA-1 can exhibit functional antagonism via direct interaction [129–132]. Therefore, early models proposed that the stochastic fluctuations in PU.1 and GATA-1 levels within early HSPCs directed the lineage choice between megakaryocytic/erythroid and lympho/myeloid fates. However, more recent studies tracking PU.1 and GATA-1 protein expression during HSPC differentiation challenged this model, instead proposing that high levels of PU.1 and GATA-1 instead occur downstream of the earliest lineage choices [133]. An early HSPC differentiation state at which PU.1 and GATA-1 proteins are co-expressed at high levels was indeed not identified in this study [133]. These findings suggest that early LMPP/EMPP branching is driven by factors upstream of PU.1 and GATA-1, which instead reinforce and drive differentiation pathways downstream of the LMPP and EMPP, respectively.

C/EBPα, a member of the CEBP family of leucine zipper TFs, is the master regulator of the myeloid lineage. C/EBPα is required for the transition from LMPP to GMP, as conditional C/EBPα deletion in mice results in the loss of GMP and granulocytes [134, 135]. Accordingly, introduction of C/EBPα in normal [136] and AML [137] cells promotes myelopoiesis. C/EBPα activity is mediated by its binding to cis-regulatory elements of several myeloid genes, such as the IL-6 receptor [136, 138]. Similarly to PU.1, C/EBPα antagonizes alternative fates, such as erythroid differentiation [139]. Whilst critical for this early stage of myelopoiesis, C/EBPα is not required downstream of the GMP [135]. In fact, the cooperative action of C/EBPα and PU.1 is thought to prime the earliest progenitors for myeloid differentiation [140]. Downstream of the GMP, TFs like C/EBPε [141], C/EBPδ, C/EBPβ [142], IRF8 [143], and GFI-1 [144, 145] regulate granulocytic and monocytic commitment.

Another important lympho-myeloid TF is RUNX1/AML1. During development, RUNX1 is critical for emergence of definitive HSCs [146]. In adult BM, RUNX1 is highly expressed in myeloid progenitors, moderately in lymphoid progenitors, and decreases during erythroid differentiation [147, 148]. Conditional RUNX1 deletion causes severe decrease in lymphoid progenitors and an increase in immature myeloid progenitors [149]. Partial RUNX1 deletion leads to accumulation of immature progenitors and negatively impacts long-term HSCs and stem cell function [150, 151], but its role in HSC maintenance is debated [152, 153]. RUNX1 haploinsufficient granulocyte progenitors commit to, but do not complete, maturation [154]. Overall, RUNX1 is required for proper differentiation of early lympho-myeloid progenitors, presumably LMPPs, but its exact role in adult BM is less precisely understood compared to PU.1 and CEBPα.

On the other side of the lympho-myeloid branch, a series of well-characterized TFs drives specification and commitment of the B-cell lineage [111]. Downstream of PU.1, IKAROS directs LMPPs towards the B cell lineage [101]. Consistently, IKAROS-null mice lack cells from even the earliest stages of B-cell development [155]. Subsequently, the TF E2A stimulates expression of EBF1 [156], a driver of early B-cell lineage specification critical for B-cell generation [111, 157]. E2A and EBF1 then induce the expression of PAX5, a TF which initiates and maintains B-cell commitment [158].

Other differentiation pathways downstream of the LMPP are less understood, possibly due to the absence of a single lineage-defining master TF regulator. One example is plasmacytoid dendritic cell (pDC) maturation. As for all lympho-myeloid paths, PU.1 and RUNX1 are required for differentiation of early pDC progenitors [159, 160]. Early pDC specification is subsequently driven by TCF4 (E2-2) [161] and BCL11A [162], and further downstream maturation is promoted by IRF8, IRF4 [163], SPIB [164], and RUNX2 [165].

Although the role of several lineage-affiliated TFs has been elucidated in normal haematopoiesis, our understanding of how these TFs get corrupted in malignant disease is limited. Some lineage-affiliated TFs, like RUNX1 and C/EBPα, are subject to recurrent somatic mutations and chromosomal aberrations in AML,

highlighting their critical role in executing proper differentiation. In fact, AML cells are most frequently arrested in differentiation at the early lympho-myeloid differentiation stages. A better understanding of how leukaemia-associated somatic lesions blunt the function of lineage-associated TFs is thus urgently required.

1.2 Malignant haematopoiesis

Somatic mutations, cytogenetic abnormalities, and epigenetic alterations disrupt normal haematopoietic differentiation, leading to blood cancer. When these aberrations occur within immature myeloid or lympho-myeloid BM cells, the uncontrolled proliferation and impaired differentiation of abnormal immature BM cells drive myeloid transformation. Several myeloid malignancies, like AML, myelodysplastic syndromes (MDS), or myeloproliferative neoplasms (MPN), are distinguished based on morphologic, immunophenotypic, cytogenetic, and genetic readouts. The focus of this thesis will primarily be on AML.

1.2.1 Acute myeloid leukaemia: an overview

AML is the most aggressive adult acute leukaemia which represents a serious unmet clinical need. AML results from differentiation arrest and clonal expansion of immature cells (blasts) within the BM [166]. The transformed cells, arising due to the acquisition of somatic lesions, gain aberrant self-renewal potential and supplant normal haematopoiesis, leading to fatal deficiencies in blood production. AML can occur in both children and adults, but it predominantly affects elderly individuals. Whilst intensive clinical and pre-clinical research led to regulatory approval of several drugs and improved the survival of younger patients, the outcome remains dismal in older patients [167], with relapse being the primary cause of mortality. Younger ($<$ 60 years) AML patients are usually treated with several cycles of intensive chemotherapy, potentially coupled to targeted therapies. Patients at higher risk of relapse undergo allogeneic stem cell transplant. Although this treatment can be curative, relapse remains commmon. For older patients, the current standard of care is the non-intensive combination of the hypomethylating agent azacitidine and

the BCL-2 inhibitor venetoclax. The median OS for these patients is 14.7 months, and most patients succumb to their disease [168, 169].

A detailed characterization of the genetic/cytogenetic landscape of AML marked a breakthrough in AML treatment [170, 171]. The identification of recurrent genetic lesions has enabled a granular genetic classification of the disease [172], with important implications for diagnosis, risk stratification [171], clinical decision-making, and therapeutic intervention. Yet, critical questions remain about how somatic alterations drive malignant growth and survival of AML cells. The following section reviews the main classes of somatic mutations in AML, the related therapeutic implications, and key unresolved challenges.

1.2.2 Genomic landscape of AML and emerging targeted therapies

Recurrent AML driver somatic lesions affect a diverse set of genes, encoding signalling molecules, haematopoietic TFs, epigenetic regulators, DNA damage response effectors, metabolic enzymes, and splicing factors [171]. Although diverse, these somatic aberrations converge to: (a) disrupt haematopoietic differentiation; (b) confer a survival advantage; and (c) enable uncontrolled proliferation. Interestingly, AML is one of the least mutated cancer types, with an average of 13 coding mutations per patient, 5 of which recurrent [173]. This has two important implications: (a) individual genetic lesions in AML exhibit potent effects; and (b) mutant cells are strongly dependent on one or few genetic alterations. Accordingly, several AML driver lesions provide unique therapeutically targetable vulnerabilities (reviewed in [174], Appendix A.1)(Figure 1.3). In some instances, this led to important breakthroughs in treatment. Here, I provide a brief overview of those targeted therapies relevant for the interpretation of the results presented in this thesis.

Activating gain-of-function (GoF) mutations in the *FLT3* gene occur in $> 30\%$ of AML patients and confer poor prognosis. *FLT3* encodes a cytokine receptor upstream of the RAS/MAPK, PI3K/AKT, and JAK/STAT pathways, which is highly expressed in LMPPs [98–100, 175, 176]. *FLT3* mutations lead to ligand-independent, constitutive activation of pro-survival and pro-proliferative pathways

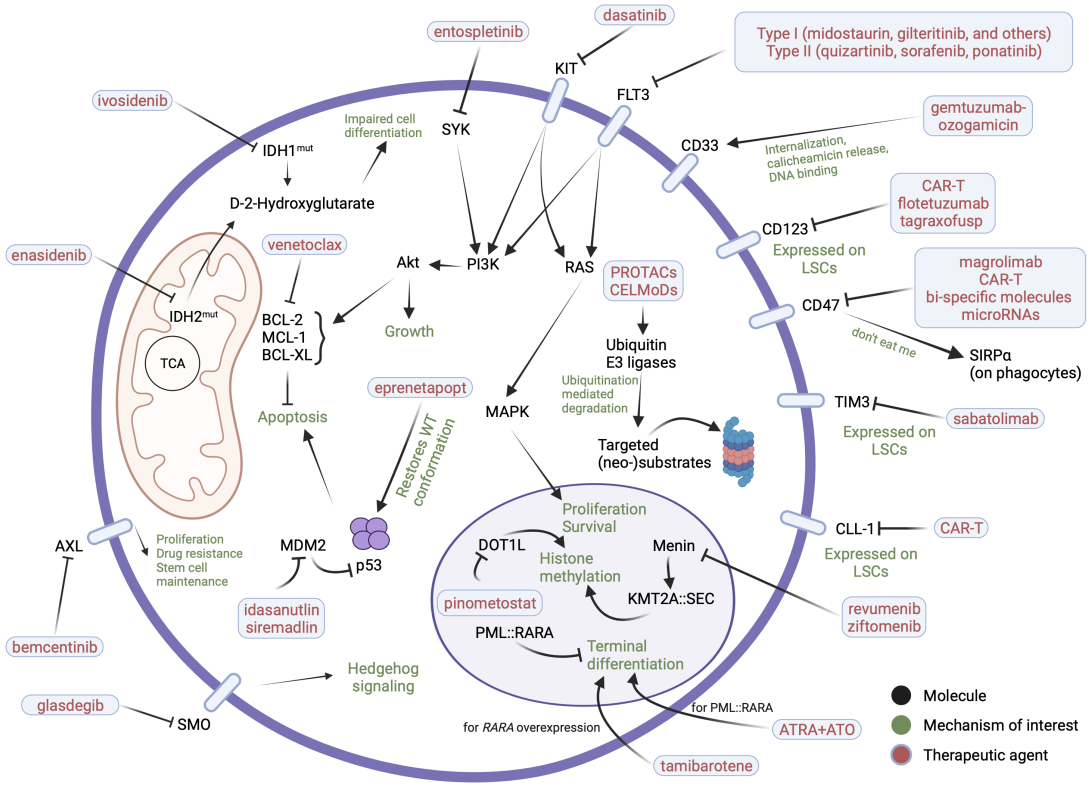


Figure 1.3: Targeted therapies in AML. Therapeutic agents are indicated in red. Some of these agents are approved for AML treatment, whilst others are undergoing clinical trials or pre-clinical evaluation. Reprinted from [174], with permission from Wiley.

[176–178]. Several inhibitors of the mutant FLT3 protein significantly improved survival of *FLT3*-mutant patients in Phase III clinical trials [179–181], which led to their regulatory approval.

GoF mutations involving the conserved residues R132 and R140/R172 in the *IDH1* and *IDH2* genes, respectively, occur in ≈20% of AML patients [171, 182, 183]. These genes encode isocitrate dehydrogenase enzymes, which normally catalyse the conversion of isocitrate to α-ketoglutarate. The mutant enzymes have neomorphic function and convert α-ketoglutarate into the oncometabolite 2-hydroxyglutare (2-HG) [184, 185]. 2-HG inhibits α-ketoglutarate-dependent enzymes, such as TET 5-methylcytosine hydroxylases and Jumonji-C domain histone demethylases [186, 187], causing aberrant DNA methylation and impaired differentiation of mutant cells [188]. Ivosidenib and enasidenib, allosteric inhibitors of mutant IDH1 and IDH2, respectively, decrease 2-HG levels and significantly improve survival of *IDH1/2*-

mutant patients [189–191]. Both drugs are now approved for this patient subgroup, and are used together with azacitidine [192, 193].

The anti-apoptotic protein BCL-2 is not recurrently mutated in AML but sustains AML cell survival [194, 195]. Consistently, the BCL-2 inhibitor venetoclax showed clinical activity in AML [196] and, when combined with azacitidine, significantly improved survival versus azacitidine monotherapy in patients ineligible for intensive chemotherapy [168]. Interestingly, venetoclax and azacitidine are particularly effective in *IDH1/2*-mutated patients [168, 197], presumably due to increased dependency of *IDH*-mutated cells on BCL-2 [198]. Based on this evidence, ivosidenib and venetoclax ± azacitidine were recently combined in *IDH1*-mutant patients in a phase Ib/II trial, which showed therapeutic benefit [199](detailed in Chapter 3).

Other mutant molecules, like epigenetic regulators or TFs, have been more difficult to target therapeutically. However, oncogenic proteins arising from fusions involving the *MLL* gene (*MLL*-rearranged, *MLL*r) exemplify efficient targeting of epigenetic regulators. Normally, the histone methyltransferase MLL1 (*KMT2A*) primes genes for activation via trimethylation of histone 3 lysine residue 4 [55]. MLL supports an active chromatin state via association with RNA polymerase II [200], maintaining transcription of several stemness-associated genes, like the *HOX* cluster [55]. However, in both infant and adult leukaemia, *MLL* can be fused to several partner genes, most frequently members of the superelongation complex. This produces potent oncogenes which abnormally promote expression of stemness-associated genes, like *HOX* family, *PBX3*, *MEF2C*, *SOCS2*, *JMJD1C*, *MEIS1*. Failure to silence these genes confers self-renewal properties to *MLL*r progenitors [201]. These effects depend on the binding of the *MLL* fusion protein to menin, encoded by *MEN1* [202]. Menin and MLL fusions coordinately regulate transcription [202–205]. Accordingly, pharmacologic disruption of this cooperation induces differentiation of *MLL*r cells [206–208]. Moreover, menin inhibitors are also active against *NPM1*-mutant AML (≈30% of AML cases) [209, 210], in which mutant NPM1 cooperates with menin and MLL [211, 212]. Menin inhibitors are showing great promise in *MLL*r and *NPM1*-mutated patients [213–216].

The success stories described above were not replicated for other recurrent AML mutations. For example, AML harboring mutations in TFs like *RUNX1*, *CEBPα*, *WT1*, or *TP53*, or in epigenetic regulators, such as *BCOR*, *EZH2*, *DNMT3A*, *TET2*, and *ASXL1*, currently lack specific targeted treatments. Yet, most of these mutations confer poor prognosis [172]. Part of the reason is that these genes, especially TFs, were traditionally considered undruggable. Moreover, our understanding of the biology of these mutations is limited, potentially due to pleiotropic and context-specific activities of these molecules. The path leading to menin inhibitors demonstrates that precise knowledge of the mechanisms downstream of mutations can guide therapeutic intervention. Hence, mechanistic understanding of the consequences of these mutations in AML is urgently required.

Despite the increasing armamentarium of AML therapies, most patients still relapse due to development of resistant disease. The key question is thus: why does treatment failure occur? Two main reasons are likely to explain most, if not all cases of treatment failure: clonal adaptation to treatment and persistence of cells with leukaemia propagating potential. As reviewed in the following sections, these two phenomena are closely intertwined.

1.2.3 The differentiation hierarchy in AML and the leukaemia stem cell

Traditionally, cancer was believed to arise from uncontrolled clonal proliferation of homogeneous malignant cells. Today, it is appreciated that cancers can instead be organized in a hierarchical fashion, partially resembling normal differentiation hierarchies. In this model, a relatively rare, primitive, self-renewing population of cancer stem cells (CSCs) gives rise to more differentiated cancer proliferating cells, which compose the tumor bulk. AML was one of the first cancers for which these concepts have been formally demonstrated, acting as a prototype of hierarchically organized malignancy. Indeed, the hierarchical organization of AML partially mirrors normal haematopoietic differentiation.

Already in 1970, rare, dormant populations of leukaemic cells were observed by 3H-thymidine infusion experiments in AML patients [217], giving rise to the

hypothesis of a putative 'leukaemic stem cell' (LSC), a self-renewing population capable of maintaining leukaemia. The subsequent development of *in vitro* colony forming assays showed that only some AML cells could generate colonies and serially replate (i.e. they had clonogenic potential). These cells were termed AML colony forming units (AML-CFU)(reviewed in [218, 219]). Colony assays revealed that AML-CFU consistently displayed surface antigens of a cell population which was less differentiated than the patients' bulk leukaemic population. AML-CFU from different patients displayed differing levels of differentiation, resembling either more primitive or more committed myeloid progenitors. Moreover, it was observed that cells within individual colonies were sometimes more differentiated compared to the original AML-CFU which formed the colony [220]. AML-CFU from most patients could serially replate, demonstrating self-renewal capacity. Replating efficiency differed between patients, and low efficiency correlated with increased remission rates [221]. Therefore, the AML-CFU has three key properties: (a) high proliferative capacity, demonstrated by the ability to form a colony; (b) self-renewing potential; and (c) differentiation potential.

The subsequent development of FACS and *in vivo* xenotransplantation (section 1.1.1) enabled the functional characterization of LSCs. The formal proof for the existence of the AML LSC was provided by the Dick group in 1994 [222]. In this seminal study, the $CD34^+CD38^-$, $CD34^+CD38^+$, and $CD34^-$ cell fractions from 1 AML patient were transplanted into severe combined immunodeficient (SCID) mice. Only the $CD34^+CD38^-$ fraction successfully engrafted and gave rise to leukaemia. It was thus proposed that the $CD34^+CD38^-$ fraction contained SCID leukaemia-initiating cells (SL-IC). In 1997, Dominique Bonnet and John Dick corroborated these findings [223]. When cells from 7 AML patients were transplanted, only the $CD34^+CD38^-$ fraction consistently engrafted. Importantly, transplanted $CD34^+CD38^-$ fractions gave rise to more differentiated $CD38^+$ and $CD33^+$ progeny *in vivo*, whilst also retaining a small proportion of $CD34^+CD38^-$ cells. Moreover, 4/4 tested samples showed efficient secondary engraftment from the $CD34^+CD38^-$ fraction, demonstrating long-term self-renewal capacity within

this fraction. Whilst confirming that SL-ICs could self-renew, this study also highlighted that differentiation in AML is not fully inhibited, with leukaemic cells being organised in a hierarchal fashion. It was thus proposed that the cell-of-origin (COO) of AML lies within the $CD34^+CD38^-$ fraction and that the SL-IC resembles the HSC. Exclusive engraftment from the $CD34^+CD38^-$ fraction was confirmed by others [224]. 3H-thymidine assays showed that, in most patients, SL-ICs were predominantly in the G_0 phase of the cell cycle, although they could spontaneously cycle in culture, consistent with SL-ICs being quiescent but harbouring cycling potential [225]. Using lentiviral vector clonal tracking, the Dick group further showed that LSCs existed in diverse functional states, with some LSCs being more competent than others in giving rise to progeny in primary and secondary recipients [226]. It was proposed that LSC diversity mirrors normal HSC heterogeneity, supporting the hypothesis that the LSCs may originate from normal HSCs.

However, this concept has since been challenged. Already in 1997, it was shown that SL-IC lacked surface expression of CD90, in contrast to normal HSCs [227]. Moreover, the same study showed that, for one AML patient, the $CD34^-$ fraction could engraft successfully, suggesting LSC heterogeneity between patients. In 2008, the Bonnet group showed that the CD38 antibody seriously ablated the engraftment capacity of leukaemic cells, presenting evidence that prior universal failure of $CD34^+CD38^+$ fractions to engraft was likely a technical artefact [228]. Since then, multiple studies demonstrated that LSC function could reside in $CD34^+CD38^-$, $CD34^+CD38^+$, and $CD34^-$ fractions, and established that AML LSC resembled early progenitors and precursors, rather than HSCs [100, 110, 228–236]. Consistently, LSCs can be generated from experimental introduction of potent oncogenes into haematopoietic progenitors [201, 237–239]. Collectively, these studies indicated that haematopoietic transit-amplifying cells which acquire self-renewal potential, and not necessarily HSCs, initiate AML.

A progenitor-like phenotype of the human LSC has been comprehensively characterized by the Vyas group in 2011 [100]. This study showed that, in most $CD34^+$ AMLs, the aberrantly expanded populations matched $CD34^+CD38^-CD90^-$

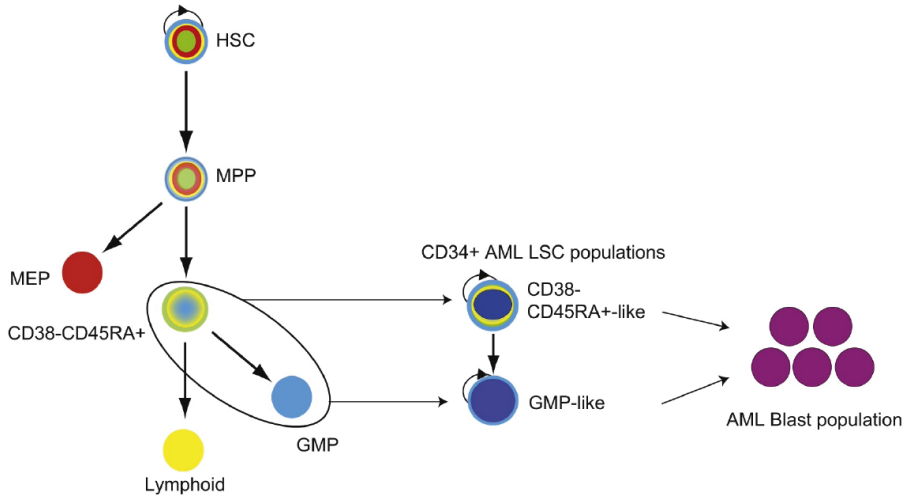


Figure 1.4: LMPP-like and GMP-like LSC hierarchy. $CD38^{-}CD45RA^{+}$ denotes immunophenotypic LMPPs. Reprinted from [100], with permission from Elsevier.

$CD45RA^{+}$ LMPPs and $CD34^{+}CD38^{+}CD123^{+/low}CD45RA^{+}$ GMPs. Transcriptionally, leukaemic LMPP-like and GMP-like cells most closely resembled normal progenitors, rather than HSCs, but upregulated stemness-associated genes, compared to normal progenitors. Both populations could engraft in mice, although the LMPP-like population had superior engrafting potential, suggesting LSCs were more frequent within the LMPP-like fraction. Importantly, leukaemic LMPP-like cells could give rise to GMP-like cells, but not vice versa, demonstrating that LMPP-like and GMP-like LSCs are organised in a hierarchical fashion, where the LMPP-like is upstream of the GMP-like population (Figure 1.4). This study provided sound evidence that, in AML, LSC activity within $CD34^{+}CD38^{-}$ BM cells was harboured within early lympho-myeloid progenitor-like cells. Of note, a smaller proportion of $CD34^{+}$ AMLs had expanded $CD34^{+}CD38^{-}CD90^{-}CD45RA^{-}$ MPP and $CD34^{+}CD38^{+}CD123^{+}CD45RA^{-}$ CMP compartments [100], later shown to also harbour LSC activity [240]. Importantly, in AML cases composed predominantly from $CD34^{-}$ cells, LSC activity resides in the $CD34^{-}$ fraction [229, 230] and, more specifically, within the $CD34^{-}CD117^{+}$ GM precursor-like fraction [110]. More recently, in a subset of patients, the Jordan group has identified LSC activity even within the $CD34^{-}CD117^{-}$ fraction, termed monocytic LSCs [234]. Overall, LSCs arise from transformation of normal progenitors, GM precursors,

and, occasionally, more differentiated myeloid cells, and sustain abnormal AML differentiation hierarchies. Of note, LSC immunophenotype varies between different myeloid malignancies: for example, LSCs in MDS resemble immunophenotypic HSCs [241–243].

In the last decade, substantial molecular characterization of human AML LSCs took place. In 2010, the Majeti group has shown that, compared to downstream leukaemic cells, LSC fractions upregulate stemness-associated genes like *HLF*, *MEF2C*, and *SETBP1* [244]. A year later, the Dick group showed that both the $CD34^+CD38^-$ and the $CD34^+CD38^+$ functionally defined LSC fractions (i.e. showing successful primary and secondary engraftment) upregulate HSC-related genes, suggesting LSCs are an abnormal hybrid between HSCs and progenitors [231]. Indeed, compared to non-LSC leukaemic cells, LSC-enriched fractions upregulate *HLF*, *SOCS2*, *MEIS1*, *MECOM*, *ERG*, *INPP4B*, and *HOX* family, highly expressed in HSC. These comparisons were further refined to generate prognostic LSC signatures [236]. Importantly, in all the above studies, high LSC gene signature scores correlated with poor patient survival, suggesting a link between LSCs and clinical outcome. Further methylation profiling indicated that functional LSCs displayed hypomethylation of several stemness-associated genes, like the *HOXA* family and *SOCS2*, suggesting HSC programs in LSCs are epigenetically encoded [235]. LSCs also display a typical metabolic state, with low levels of radical oxygen species and high rates of mitophagy [195, 245]. It has also been proposed that LSCs, at least those of certain genotypes, upregulate signalling pathways like WNT/β-catenin [246, 247], Sonic Hedgehog [248, 249], or NF kB [250]. However, these concepts have still not been translated into efficient, long-lasting LSC-targeting therapies, potentially due to the heterogeneity of the LSC pool.

More recently, single-cell RNA-seq (scRNA-seq) technologies corroborated that AMLs are organized in an abnormal continuum of differentiation, partially resembling normal myeloid and/or lympho/myeloid differentiation [88, 91, 251–253]. Within the hierarchy, malignant cells can either reside within several differentiation states or may predominantly exist at one progenitor-like stage. Such heterogeneity

has therapeutic implications, as hierarchy composition correlates with survival and response to therapy [251, 254]. However, gene expression within the AML hierarchy is not equivalent to that in normal differentiation, likely reflecting inheritance of disordered LSC programs. For example, HSC and GMP signatures are frequently coupled in AML cells.

In summary, molecular and biochemical studies confirmed that LSCs possess hybrid characteristics between early progenitor and primitive stem cell states.

1.2.4 The role of LSCs in treatment failure

The early LSC studies already hypothesized that persistence of quiescent, therapy-resistant LSCs may correlate with treatment failure [221, 224, 225, 255], supported by the predictive value of LSC scores [231, 236, 244, 256]. The current understanding is that the high relapse rates in AML, following initial clinical response, are due to failed eradication of LSCs. Evidence for this hypothesis was provided by the correlation between the rates of immunophenotypic LMPPs and GMPs at morphological complete remission (CR) and incidence of relapse, in patients treated with azacitidine and sodium valproate, or with chemotherapy [257]. In 5 patients who achieved CR, immunophenotypic LMPP-like and GMP-like compartments at CR were at least 6-fold larger than in normal BM, despite the absence of frank leukaemic expansion. Consistently, all 5 patients relapsed, and clinically frank increase in blast counts could be preceded by LMPP/GMP expansion. Meanwhile, LMPP/GMP reached normal size at CR in 3 patients who achieved durable remission. Hence, failed eradication of immunophenotypic LSCs during clinical remission presaged relapse. In this regard, LMPP and GMP populations likely harbour a fraction of more dormant, therapy-resistant cells, which reversibly enter cell cycle and give rise to relapse, acting as *bona fide* LSCs [258]. Similarly, in patients achieving clinical response after treatment with azacitidine and vorinostat, the frequency of LMPPs in the BM at remission often remained above normal physiological levels. In some of these patients, subsequent LMPP re-expansion coincided with occurrence of relapse [259]. LSCs can increase in numbers and change immunophenotypically at relapse following

induction chemotherapy, suggesting therapeutic pressure stimulates symmetric LSC divisions [233]. Persistence of LSC at remission is related to measurable residual disease (MRD), a state in which a proportion of leukaemic cells is still detectable in the BM, despite achievement of clinical response. Persistence of MRD at remission, evaluated via FACS or genetic markers, predicts poor prognosis [260–263].

Various studies were designed to identify potential LSC-specific surface markers (reviewed in [264]). Some of the best-known are TIM3, which activates both NF kB and β-catenin signalling [265–267], CD47 [268], and CD96 [269]. However, clinical trials testing treatments targeted against these markers showed mixed results, and there is still no approved drug targeting an LSC-specific marker. Importantly, the Jordan group showed that functionally defined LSCs overexpress and rely on BCL-2 for their survival [195], which may be related to the activity and safety profile of venetoclax [168, 196].

Although the functional and molecular properties of LSCs have been extensively studied, our understanding of how somatic mutations gradually convert normal progenitors into LSCs and how cooperative somatic lesions support LSC phenotypes is limited. Yet, unravelling these processes will hold the key towards uncovering LSC vulnerabilities and reducing relapse rates.

1.2.5 The clonal hierarchy in AML

It is increasingly appreciated that LSCs at relapse are often not genetically equivalent to LSCs at diagnosis. In fact, besides the AML differentiation hierarchy, a substantial source of AML heterogeneity arises from the genetic clonal hierarchy. Whilst the differentiation hierarchy consists of several malignant differentiation states, the clonal hierarchy is composed from multiple genetic clones. These two hierarchies are intertwined: a genetic clone can co-exist at multiple differentiation stages, and a single differentiation stage can be polyclonal. When a mutation, or a combination of mutations, disrupts differentiation capacity, the resulting genetic clone is arrested at one or more immature differentiation stages [251, 270–276] (Figure 1.5).

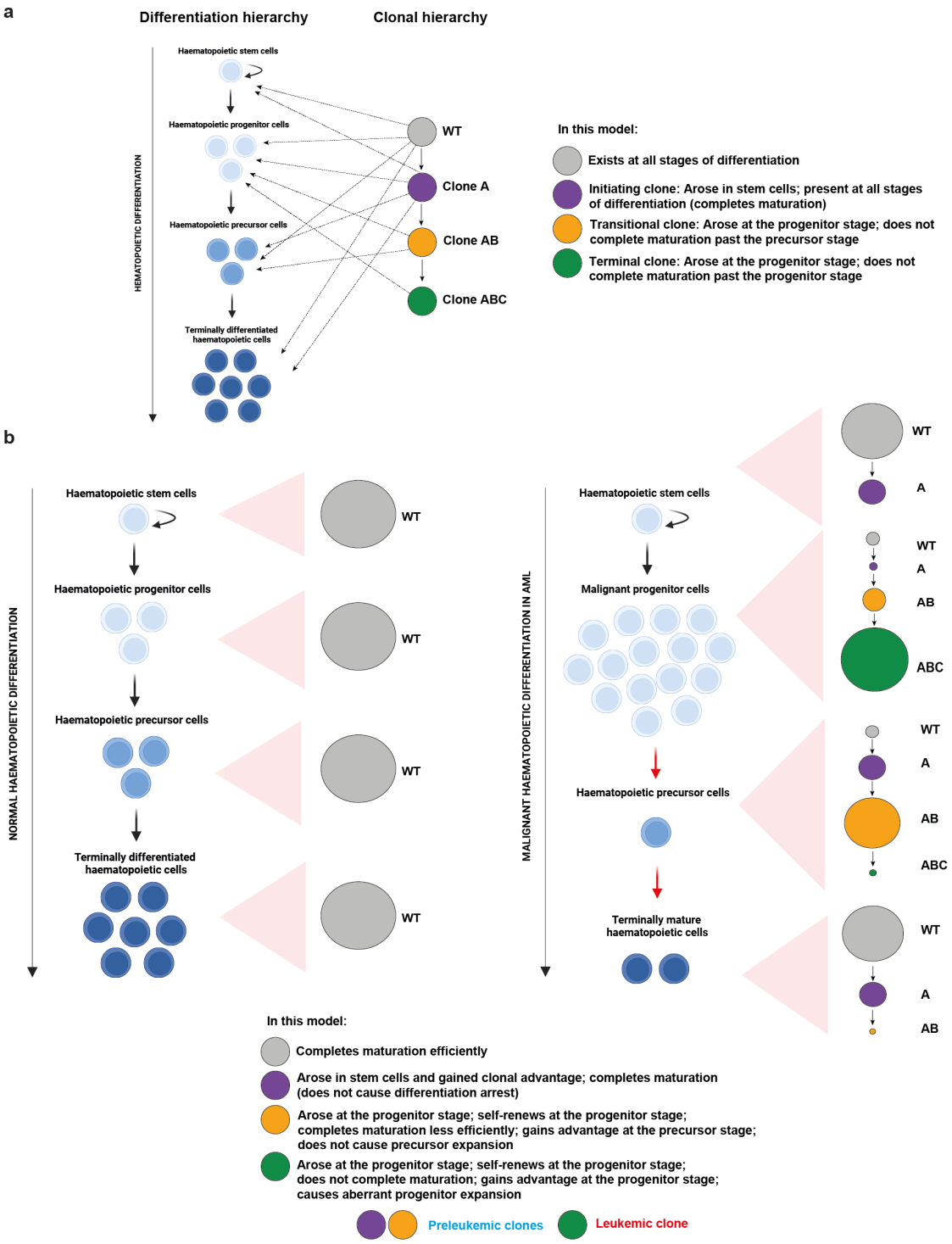


Figure 1.5: Link between clonal and differentiation hierarchies in AML. (a) Exemplar model for the link between the differentiation (left) and clonal (right) hierarchies in AML. The legend summarizes differentiation dynamics of each clone in this model. **(b)** Normal (left) and malignant (right) differentiation. In a normal BM, where cells do not harbour driver somatic lesions (WT), self-renewing HSCs differentiate through progenitors and precursors to terminally mature blood cells. In AML (right), multiple genetic clones co-exist within haematopoietic populations. The circle size (right) represents the fraction of each clone at each differentiation stage. The legend explains the dynamics of each clone in this particular scenario.

Large next generation sequencing (NGS) efforts have established that the pathogenesis of adult AML is, in most cases, a stepwise progress, originating from gradual acquisition of driver somatic lesions. AML is a clonal disease, meaning that all malignant cells within a patient derive from a single initial 'founder' COO. In this model, a single, long-lived cell acquires a somatic driver mutation which confers clonal advantage, leading to clonal expansion. However, this 'pre-leukaemic' mutation is not sufficient to cause leukaemia. Further 'transforming' driver mutations occur subsequently within the pre-leukaemic clone, disrupting normal differentiation and leading to frank AML.

Seminal evidence for existence of pre-leukaemic clones came in 2000, from the observation of AML patients who were in durable remission [277]. The *AML1/ETO* leukaemic fusion transcript was detected in a proportion of immunophenotypic HSCs, CLPs, monocytes, and B cells at remission, even up to 150 months after treatment initiation. This suggested that a proportion of non-malignant, persistent, self-renewing HSCs contained the fusion and differentiated normally. Hence, the fusion alone was not sufficient to drive leukaemia. The concept of pre-leukaemic mutations was definitively established in 2012 [278]. In this study, residual, non-malignant, immunophenotypically normal HSCs/early progenitors were isolated from AML patients. In 5/6 patients, some, but not all AML mutations, were present in non-malignant cells. Mutations present exclusively in AML cells, frequently affecting *NPM1* and *FLT3*, were thus late, transforming events. The residual normal HSCs/early progenitors were termed pre-leukaemic, and it was proposed that pre-leukaemic cells were the predecessors of fully transformed, malignant populations. In fact, pre-leukaemic cells can complete terminal differentiation (measured by the presence of the pre-leukaemic mutations in terminally mature cells) and persist at remission [279]. Moreover, it was observed that pre-leukaemic mutations usually occur in global epigenetic regulators, whilst later mutations occur in genes encoding pro-proliferative signalling molecules, myeloid TFs, and *NPM1*, conferring strong proliferative/survival advantage or differentiation arrest. In a parallel study [280], it was shown that residual normal immunophenotypic

HSCs from *DNMT3A*/*NPM1*-mutant AML patients harboured *DNMT3A*, but not *NPM1* mutations, suggesting *DNMT3A* mutations occurred prior to *NPM1* mutations, the latter always present in leukaemic blasts. Residual normal HSCs from diagnostic samples sustained primary multi-lineage engraftment, indicating that pre-leukaemic cells functionally resembled normal rather than leukaemic cells. However, whereas the initial *DNMT3A* variant allele frequency (VAF) in the engrafted mice was 20%-30%, indicating these residual HSCs were a mixture of wild type (WT) and *DNMT3A*-mutant cells, the *DNMT3A* VAF increased over time, indicating *DNMT3A*-mutant HSCs had competitive advantage.

In the same year, three studies observed an age-related clonal expansion of blood cells harbouring an AML-associated pre-leukaemic driver mutation in individuals without haematologic malignancy [281–283]. This phenomenon was termed clonal haematopoiesis (CH). Strikingly, mutations in three epigenetic regulators, namely *DNMT3A*, *TET2*, and *ASXL1* (abbreviated DTA), were the most frequent somatic events. Other common CH mutations include those in splicing factors, whilst there is paucity of mutations in TFs, like *RUNX1*, *NPM1* mutations, or mutations affecting proliferative genes. In agreement with the CH clone potentially being pre-leukaemic, CH is associated with ≈10-fold increased risk of hematologic malignancy [281]. To identify potential strategies for prevention or delay of AML, a vast body of work has since been devoted to understanding how CH mutations fuel clonal outgrowth. Studies adopting genetic modification of mouse or human cells followed by *in vitro* assays and xenotransplantation recapitulated clonal advantage of DTA-perturbed cells [64–66, 69, 284, 285] and suggested clonal advantage arises either due to increased expression of self-renewal genes [69, 286, 287], or relative resistance of CH clones to inflammation [288–292], the latter being recently proposed in human CH [80](Appendix A.2).

CH mutations cooperate with subsequent transforming lesions to drive leukaemic transformation [209, 271, 273, 275, 293–295], converting immature BM cells into LSCs. The transforming potential of *FLT3* or *NPM1* mutations in synergy with pre-leukaemic mutations has been well documented [209, 293–295]. Yet, how other late

mutations, like those in TFs, cooperate with pre-leukaemic mutations to drive AML, has been more challenging to disentangle. As CH transforms to AML, the number of mutations, but even more so the number of genetic subclones, increases [275], driven by linear or branched clonal evolution [270, 274, 275]. Of note, clonal hierarchies are not established randomly: whilst some mutation combinations are more frequent due to mutational synergy, other mutations are mutually exclusive due to functional redundancy [171, 273–275]. For example, *RUNX1* and *BCOR* mutations frequently co-occur in AML [296–299], but the mechanistic basis for this is unclear.

The substantial clonal heterogeneity in AML likely accounts for poor patient outcome. First, there is notable heterogeneity in driver mutations observed between patients, and only 3 mutations (*DNMT3A*, *NPM1*, and *FLT3*) occur in $> 25\%$ of AML patients [170]. Second, the same mutation can co-occur in different clonal contexts across patients, which likely affects its downstream consequences. Third, clonal heterogeneity has phenotypic relevance, as distinct mutations preferentially accumulate at distinct stages of AML differentiation [110, 251, 254]. Therefore, studying the effects of mutations in the appropriate clonal and cellular context and understanding mutation cooperativity are crucial to unravel clone-specific vulnerabilities.

1.2.6 Clonal heterogeneity and AML relapse

Residual therapy-resistant LSCs and the intra-patient clonal complexity drive AML relapse. This section will review the tight link between these phenomena.

Despite the successes of targeted therapies (section 1.2.2), treatment failure is still common. This has primarily been attributed to selection of somatic mutations which confer resistance to treatment (acquired genetic resistance). For example, secondary *FLT3* mutations [300] or RAS/MAPK mutations [301] cause resistance to FLT3 inhibition. Mutant IDH1 inhibition can be bypassed by mutations affecting epigenetic regulators and TFs [270, 302], which re-establish differentiation block, by *FLT3* and RAS/MAPK mutations [270, 302], and by either second site *IDH* mutations [303] or those affecting the alternative IDH isoform [304]. Genetic

resistance to venetoclax also arises due to mutations *FLT3* or RAS/MAPK pathway [305], which upregulate alternative anti-apoptotic proteins [306], but also *TP53* mutations [307]. In ≈40% of patients with acquired resistance to Menin inhibition, *MEN1* gene mutations abolish inhibitor binding to menin [308].

Hence, genetic clones which adapt to treatment expand and lead to disease relapse. Accordingly, whole-genome sequencing of diagnostic and relapsed samples from chemotherapy-treated AML patients revealed that acquisition of resistance-associated mutations, either in the founder pre-leukaemic clone or within a leukaemic subclone, associates with relapse [272]. Thus, in most cases, evidence suggests relapse arises from incomplete eradication of diagnostic clones, which evolve to resist treatment, rather than from emergence of completely independent novel clones, as supported by other studies [263, 309, 310]. More recently, patterns of clonal evolution underpinning relapse have been formally demonstrated by the advent of single-cell genotyping [199, 270, 274, 275, 311, 312]. Whilst these studies confirmed that relapse-associated clones are genetically linked to diagnostic clones, they also revealed a remarkable degree of clonal diversity underpinning resistance to different therapies. This highlights that each therapeutic regimen acts as a unique selective pressure, favouring the Darwinian outgrowth of the fittest clones.

Acquired genetic resistance is linked to the AML differentiation hierarchy. For the resistant clone to re-establish the disease at relapse, clonal selection needs to occur within the BM compartment harbouring LSC potential. Formal evidence that clonal selection within LSCs underpins relapse was provided in 2017 [313]. In this study, bulk genotyping of paired diagnostic and relapse patient samples was combined with xenotransplantation and genotyping of diagnostic samples. Firstly, this identified genetic variants which expanded at relapse. In 3 patients, relapse-associated variants were not detectable within diagnostic AML blasts, but were present at low VAF within a primitive LMPP-like population. This indicated that, albeit not dominant at diagnosis, relapse-associated clones existed already pre-treatment within a subset of (presumably dormant) LSCs. Therapy-induced clonal selection within LSCs gave rise to relapse. Accordingly, xenotransplantation

of diagnostic samples produced leukaemic grafts with detectable relapse-associated variants (i.e. mutant LSCs were experimentally induced to proliferate). In 4 other patients, relapse variants were detectable within more differentiated diagnostic AML blasts. This suggested that the relapse-originating LSCs may reside within both $CD34^+$ and $CD34^-$ populations, consistent with prior evidence [100, 110, 234]. Clonal selection thus occurs within LSCs, corroborating the observations that failed eradication of LSCs underpins relapse [257, 259], and showing that clonal re-organization within residual LSCs drives treatment failure.

Multi-modal single-cell technologies further demonstrated the link between clonal and differentiation AML hierarchies. Paired single-cell RNA-seq with genotyping showed that different genetic subclones tend to segregate into distinct AML differentiation stages [251, 252]. Accordingly, different somatic lesions correlate with distinct AML hierarchy compositions. For example, *RUNX1* mutations were found either within hierarchies composed by early HSPC-like cells, or by a continuum of progenitors and more differentiated cell types. These studies indicate that different somatic lesions: (a) confer context-specific clonal advantage; and (b) that different mutation combinations likely generate LSCs from distinct differentiation stages. Accordingly, transcriptional changes accompanying AML progression are significantly correlated to the underlying clonal changes [312]. However, transcriptional adaptation to treatment can also occur in absence of genetic evolution. This suggests that clones can adapt to treatments via non-genetic mechanisms [312], reflected in the fact that ≈40% of AML relapses occur without apparent genetic evolution [314]. In these cases, dynamic changes in chromatin landscapes accompany relapse [314], suggesting epigenetic evolution as an alternative route towards treatment failure. Of note, it cannot be excluded that, in at least some of these cases, unknown driver somatic lesions also contribute to resistance.

Although the above studies advance our understanding of AML relapse, important questions persist. First, longitudinal kinetics of clonal selection associated with resistance are not well understood. Specifically, how rapidly post-treatment are resistant clones selected, and in which BM compartments selection operates, is not

comprehensively documented across different regimens. Yet, this knowledge may substantially improve clinical management of relapse. Second, how do mutations cooperatively alter regulation of stemness-associated genes to establish resistant LSC phenotypes remains a critical question with therapeutic implications. Finally, epigenetic mechanisms of clonal adaptation to treatment, likely accounting for a substantial proportion of relapses, remain underexplored, due to the paucity of methods allowing to link clonal and epigenetic readouts in primary patient samples.

1.2.7 Model systems to study the consequences of somatic mutations

Targeted genetic intervention in model systems has elucidated how somatic mutations alter cellular phenotypes. However, knowledge of how cooperating mutations sustain malignant clones, necessary for the development of personalised treatment options, remains challenging to study. In this regard, mouse models harbouring multiple mutations have proven useful [209, 295, 315–318]. Such models have revealed synergistic dysregulation of DNA methylation, TF activity, and gene expression, downstream of *FLT3*-ITD mutations in *TET2*-, *DNMT3A*-, or *NPM1*-mutant background [295, 316, 318]. Importantly, these pre-clinical models exemplified how malignant cells harbouring multiple mutations can concurrently have specific therapeutic vulnerabilities [209, 315]. More recently, CRISPR/Cas9 editing applied to primary human $CD34^+$ cells enabled modelling of cooperative mutations in human settings. For example, this revealed that *BCOR* mutations can rescue stem cell dysfunction in *RUNX1*-deficient cells, potentially explaining their frequent co-occurrence [319]. Albeit powerful, these models do not accurately recapitulate the stepwise acquisition of somatic lesions in human AML. To begin to address this, the Levine group has developed a mouse model where *DNMT3A*, *NPM1*, and *FLT3* mutations can be induced sequentially, by using different driver recombinases [294].

Nevertheless, the above model systems may still not fully capture the complexity of human disease. First, in human AML, multiple distinct genetic subclones co-exist and vary dynamically throughout treatment, which is challenging to model. Second, in different patients, somatic lesions can occur in differing COOs, which

may translate into differential functional and molecular properties associated with the same somatic lesion [320]. Third, human AML develops gradually over time, under the influence of heterogeneous external stimuli, whilst mice live in controlled conditions and are exposed to a limited number of environmental pressures.

1.2.8 Linkage of genetic and molecular landscapes in primary human samples

A strategy to mitigate the limitations above is to study cancer evolution in the primary human setting. This approach is complicated by the fact that cancers are often composed from an admixture of genetic clones, which exist in complex clonal and differentiation hierarchies; clone-specific analyses are therefore required. However, prospective purification of distinct genetic clones is usually not feasible, due to the paucity of known clone-specific surface markers, thus preventing bulk comparisons between clones. In this regard, recent approaches providing simultaneous clonal and molecular readouts from the same single cells have revolutionised the study of cancer evolution. Here, a brief review of these technologies is provided, highlighting advantages and limitations of current methods.

G&T-seq is one of the first methods coupling genomic and transcriptomic profiling of single cells [321]. Here, RNA is separated from DNA by using an oligo(dT) probe annealing to the mRNA polyA tail, enabling whole transcriptome and whole genome profiling from a limited number of single cells. The copy number alteration (CNA) status was also inferred at single-cell resolution from RNA expression, thus linking genetic CNA clones to differentiation hierarchy within oligodendroglioma [322]. Further studies on glioma adopted the scRNA-seq chemistry termed Smart-Seq2, to infer CNA status and read single nucleotide variants (SNVs) directly from RNA [323]. This showed that genetic clones could segregate to distinct tumor differentiation states. However, whilst powerful in CNA-rich cancers, these approaches had limited usage in malignancies driven largely by point mutations or small insertions/deletions (like AML), due to sparse SNV coverage.

Several methods mitigated single-cell genotyping data sparsity by adopting targeted amplification of the mutant locus from mRNA. In 2017, targeted genotyping

of the *BCR-ABL* fusion gene was coupled to scRNA-seq [324]. This was achieved via the addition of *BCR-ABL*-specific primers to the reverse transcription and PCR (RT-PCR) step, which normally amplifies single-cell transcriptomes. This method discerned transcriptional differences between *BCR-ABL*+ LSCs and WT stem cells from the same *BCR-ABL*-positive CML patients. Albeit highly sensitive, this approach suffered from a low cell throughput. Improvements in this direction were achieved by adopting the nanowell technology, in which target-specific probes were used to amplify mutant loci from mRNA-derived complementary DNA (cDNA)[251]. Genotyping several driver mutations in single AML cells identified that different genetic clones segregated to distinct stages of AML differentiation. Here, a major limitation was that single-cell genotyping efficiency strongly depended on the proximity of the mutation to the 3' cDNA end. This caveat was partially resolved by the development of Genotyping of Transcriptomes (GoT), based on the 10X Chromium platform for droplet-based scRNA-seq, which leveraged target-specific primers to amplify mutant loci from mRNA [325]. Application of GoT to *CALR*-mutated MPN revealed that, although mutant and WT cells shared similar differentiation trajectories, *CALR* mutations biased cells towards specific fates. Crucially, molecular consequences of the *CALR* mutation were not equivalent between different cell populations, exemplifying the important concept that driver mutations alter phenotypes in a differentiation stage-specific manner [325]. mRNA-based single-cell genotyping was also integrated to transcriptomic and DNA methylation profiling from the same cells [326], revealing how *DNMT3A*-R882 mutations disrupt DNA methylation in human CH [327]. Subsequently, GoT was further improved to profile genotype-specific surface proteomes and differential mRNA splicing [328].

All the above methods rely on mRNA expression for single-cell genotyping, which has two issues: (a) lowly expressed genes or those subject to mRNA decay are challenging to genotype; and (b) high rates of allelic drop-out (ADO) lead to many heterozygous mutant cells being wrongly classified as WT. Both caveats were resolved by the development of TARGET-seq, a plate-based single-cell method coupling whole transcriptome preparation with targeted genotyping from both mRNA and

genomic DNA (gDNA)[329, 330]. TARGET-seq revealed that heterogeneous genetic clones within MPN patients correlated with distinct gene expression profiles and elucidated the transcriptional basis for clonal advantage in *TP53*-mutated AML [331] and *DNMT3A*- and *TET2*-mutated CH [80](Appendix A.2). A further advance was introduced by MutaSeq, which coupled scRNA-seq to genotyping of both nuclear and mitochondrial mutations, refining clonal dissection and allowing identification of genetic clones even when mutated nuclear loci are not known a priori [253]. Application of MutaSeq to AML mapped clonal progression to changes in gene expression, revealing major molecular perturbations as genetic pre-leukaemic cells acquired transforming, leukaemic mutations. More recently, CloneTracer improved coverage of both SNVs and mitochondrial mutations in 10X Chromium, allowing high-throughput single-cell linkage of clonal and transcriptomic readouts [252].

Altogether, single-cell multi-omic approaches demonstrated a clear correlation between somatic genetic variation and altered molecular states across human myeloid disorders. The next step is now to understand the mechanistic link between gradual acquisition of cooperative somatic lesions and establishment of LSC phenotypes. By identifying the key molecular drivers of LSC maintenance downstream of diverse genetic alterations, the field may move towards the design of therapeutic strategies capable of eradicating relapse-initiating LSCs.

1.3 Aims

To improve the clinical management of AML, delay the onset of relapse, and understand how malignant phenotypes are established in patients, two broad questions need to be addressed:

1. How rapidly and in which cell populations are clonal foundations of treatment failure established? Can selection of relapse-associated clones within LSCs be detected early, to be able to predict and potentially prevent clinically overt relapse?

2. How do somatic mutations cooperatively alter gene regulation to reprogram normal progenitors/precursors into AML LSCs? How is the activity of lineage-affiliated TFs corrupted by mutations, leading to impaired haematopoietic differentiation?

To begin to address the first question, the clonal and cellular bases of either acquired resistance or sustained response to novel ivosidenib combination therapies were studied, using longitudinal samples from patients enrolled in an ongoing clinical trial. By performing this study, I reasoned that the results, although obtained from a patient cohort undergoing a specific targeted treatment, may provide broader fundamental concepts of clonal selection underpinning relapse, possibly generalizable across treatments. These analyses are presented in Chapter 3. To start addressing the second question, I reasoned that a method enabling simultaneous capture of clonal identity and chromatin accessibility at single cell resolution needs to be developed. Albeit powerful, the approaches reviewed in section 1.2.8 are still limited in their ability to generate mechanistic links between somatic mutagenesis and malignant phenotypes, because existing strategies do not capture epigenetic variation downstream of mutations. Yet, obtaining this level of information would move the field a step closer to understanding how somatic lesions reprogram normal HSPCs to become LSCs. The development and benchmarking of this method, termed GTAC, is presented in Chapter 4. The subsequent application of GTAC to human AML is presented in Chapter 5.

Chapter 2

Materials and Methods

Contents

2.1 Human samples and ethical approval 34
2.2 Outcome parameters 35
2.3 Cell culture 35
2.4 Bulk DNA sequencing 36
2.5 Primary sample staining for flow cytometry 37
2.6 Bulk cell sorting of primary human samples 38
2.7 TARGET-seq+ 38
2.7.1 TARGET-seq+: primer design and validation 38
2.7.2 TARGET-seq+: lysis buffer preparation, single-cell sorting, and RT-PCR 39
2.7.3 TARGET-seq+: 3' biased whole transcriptome library construction 41
2.7.4 TARGET-seq+: single-cell genotyping library construction 42
2.8 TARGET-seq+ analysis 43
2.8.1 TARGET-seq+: single-cell genotyping data pre-processing 43
2.8.2 TARGET-seq+: mutation calling 43
2.8.3 TARGET-seq+: FACS index data 45
2.8.4 TARGET-seq+: transcriptome data pre-processing and quality filtering 45
2.8.5 TARGET-seq+: copy number alteration inference . . . 46
2.8.6 TARGET-seq+: inference of clonal hierarchies 48
2.8.7 TARGET-seq+: inference of clone size in Lin^- cells . . 49
2.9 Genotyping integration to bulk ATAC 49
2.10 GTAC 51
2.10.1 GTAC: primer design and validation 51
2.10.2 GTAC: experimental procedure 52
2.11 GTAC analysis 54
2.11.1 Single-cell genotyping analysis 54
2.11.2 GTAC: ADO estimation 55
2.11.3 GTAC: scATAC-seq data pre-processing 56
2.11.4 GTAC: quality filtering 57
2.11.5 GTAC: dimensionality reduction, clustering, and cluster annotation 57
2.11.6 GTAC: peak calling, marker peak identification, and co-accessibility 58
2.11.7 GTAC: differential accessibility analysis 59
2.11.8 GTAC: projection on reference 59
2.11.9 GTAC: trajectory analysis 60

2.11.10 GTAC: identification of putative PRC1.1 direct targets 60
2.11.11 GTAC: TF footprinting analysis 61

Detailed TARGET-seq+ and GTAC procedures were published in prior works, where I contributed as joint first author ([80, 332, 333] and Appendices A.2 A.3 A.4). For detailed TARGET-seq+ methods (Chapter 3), please refer to Appendix A.2 [80]. For detailed methods regarding the GTAC protocol, (Chapters 4-5), please refer to Appendices A.3 [332] (original paper) and A.4 [333] (detailed STAR Protocol). Here, I provide a more concise overview of TARGET-seq+ and GTAC, as well as detailed methods regarding procedures not listed in the respective appendices.

2.1 Human samples and ethical approval

For all patient samples, written informed consent was secured from study participants in accordance with the Declaration of Helsinki. All studies involving human participants were approved by the local ethics committees at the participating centers [80, 199, 332].

Bone marrow (BM) samples from 9 patients with *IDH1*-mutated myeloid malignancies (Chapter 3) were collected at the MD Anderson Cancer Center, Houston, as part of the phase I/IIb clinical trial evaluating the combination of ivosidenib and venetoclax with or without azacitidine (NCT03471260). The patients' clinical characteristics, the treatment course, and dosages are described in section 3.2.1. Samples were collected from the following time points: pre-treatment (baseline); an early treatment response (end of cycle 1 or 3); and either relapse or longer-term remission (end of cycle 9 or 11). Karyotyping was performed as described previously, using standard metaphase cytogenetics [199]. *IDH1* mutation assessment was performed via a targeted NGS panel (section 2.4) and droplet-digital PCR (ddPCR) [199].

For the AML patient sample used in Chapter 5, cytogenetic analyses, isolation of BM mononuclear cells (MNCs), and gDNA extraction were performed as described previously [80] (Appendices A.2 A.4).

For the control WT BM used across Chapters 3-5, samples were collected from 10 individuals with no evidence of current or prior haematologic malignancy, under the Mechanisms of Age-Related Clonal Haematopoiesis (MARCH) Study, as described previously [80] (Appendix A.2).

2.2 Outcome parameters

For patients enrolled in the NCT03471260 clinical trial (Chapter 3), remission was defined according to the 2022 European LeukemiaNet (ELN) recommendations [172]. Overall survival (OS) was defined from the first day of therapy to death or last follow-up. Duration of response (DOR) was defined as the time from the first documented response to therapy, which was either blast reduction, partial response (PR), CR, or CR with incomplete haematologic recovery (CRi), to treatment failure (defined as relapse, disease progression, or the requirement to switch therapy).

2.3 Cell culture

K562, MOLM-13, and THP1 human cell lines were obtained from the Vyas laboratory stocks. The JURKAT human cell line was a kind gift from Prof. James O. J. Davies. The OCI-AML3 human cell line was a kind gift from Ms Vassilena Sharlandjieva and Prof. Thomas A. Milne. OCI-AML3 cells were cultured in α-MEM medium containing nucleosides (Cat#12571063, Gibco), 20% fetal bovine serum (FBS; Cat#F9665-100ML, Sigma), and 1% V/V Pen-Strep (Cat#15140122, Gibco). K562, MOLM-13, THP1, and JURKAT cells were cultured in RPMI-1640 medium (Cat#21875034, Gibco) containing 10% FBS and 1% V/V Pen-Strep. Prior to experiment conduction, cells were tested for Mycoplasma contamination with the MycoAlert Mycoplasma Detection Kit (Cat#LT07-218). Cells were passaged every 2-3 days and seeded aiming for a density of ≈500,000 cells/mL. Cells were maintained in a 5% CO_2 incubator at 37°C and 95% humidity. Prior to GTAC experiments, viability was evaluated using a NucleoCounter NC-3000 (ChemoMetec) to assure $>$ 90% viability.

2.4 Bulk DNA sequencing

For patients enrolled in the NCT03471260 trial (Chapter 3), an NGS-based targeted panel, which interrogates 81 genes subject to recurrent myeloid somatic mutations, at sensitivity ≈1-2%, was performed at the MD Anderson Cancer Center, Houston, as described previously [199]. A subset of patient samples was further screened by Ms. Marlen Metzner, using an extended 97-gene targeted NGS panel [80] (Appendix A.2) at a mean read depth of 822× (for pts 4/9/15).

Additionally, a subset of patient samples (pts 4/5/10/11 at baseline; pts 11/14 at relapse) was screened by Dr. Bilyana Stoilova, using whole exome sequencing (WES). Briefly, $CD45^{dim}$ blasts and $CD3^{+}$ control T-cells were sorted using a BD Fusion. gDNA was extracted using DNA Mini or Micro Kits (Cat#51304 and Cat#56304, Qiagen). WES was performed by Novogene Europe using low input library preparation method and SureSelect Human All Exon V6 bait design (Agilent). Libraries were sequenced using 150 bp paired-end reads, 50× for T cells and 100× for leukaemic blasts. Data was mapped using bwa (bwa 0.7.17) [334]. Variant calling was performed using Mutect2 (gatk 4.0.5.1) [335], Strelka2 (strelka 2.9.2) [336] and VarDict (verdict = 2018.10.18) [337]. Variants were filtered for those called by at least 2/3 callers and displaying VAF > 0.05. *FLT3*-ITD was additionally called using Pindel (pindel 0.2.5) [338]. Mutations identified by either the targeted panel or WES in all analysed samples are in Table A.1 (baseline) and A.2 (relapse).

For the healthy WT control BM (i.e. a BM sample not harbouring any of the screened mutations; Chapters 3-5) and the AML patient sample analysed by GTAC (Chapter 5), a 97-gene NGS-based targeted panel was performed by Ms. Marlen Metzner at the University of Oxford, as described previously [80] A.2. Sequencing data pre-processing, variant calling, and variant filtering were performed as described previously [80, 332].

2.5 Primary sample staining for flow cytometry

All human samples used for this thesis were previously subject to Ficoll gradient separation in order to isolate mononuclear cells, followed by cryopreservation in FBS containing 10% dimethyl sulfoxide (DMSO). Samples were stored in liquid nitrogen. Prior to TARGET-seq+ or GTAC procedures, primary human BM or peripheral blood (PB) samples were first thawed and stained for flow cytometry, as detailed previously [80, 332, 333] (Appendices A.2 A.3 A.4). Most of the procedure was equivalent between TARGET-seq+ and GTAC.

Briefly, cryopreserved samples were thawed by gradual addition of FBS and thawing medium, consisting of IMDM (Cat#21056023, Gibco), 20% FBS, and 110 μg/mL of DNase I (Cat#11284932001, Roche). Cells were centrifuged, resuspended in FACS buffer (IMDM with 10% FBS and 10 μg/mL DNase I), and filtered. Cell numbers and viability were recorded on a NucleoCounter NC-3000. For TARGET-seq+, cells were stained with the following antibodies: anti-CD38-BV421 (1:20, Cat#356618, BioLegend, clone HIT2), anti-CD10-BV605 (1:40, Cat#312222, BioLegend, clone HI10a), anti-CD117-PE/Dazzle594 (1:80, Cat#313238, BioLegend, clone 104D2), anti-CD45RA-BB515 (1:40, Cat#564552, BD, clone HI100), anti-CD123-PE (1:40, Cat#306006, BioLegend, clone 6H6), anti-CD90-PE/Cy7 (1:20, Cat#328124, BioLegend, clone HI100), anti-CD34-APC (1:160, Cat#343510, BioLegend, clone 581), using a lineage (Lin) cocktail containing: anti-CD2-PE/Cy5 (1:160, Cat#300210, BioLegend, clone RPA-2.10), anti-CD3-PE/Cy5 (1:320, Cat#300310, BioLegend, clone HIT3a), anti-CD4-PE/Cy5 (1:160, Cat#300510, BioLegend, clone RPA-T4), anti-CD8a-PE/Cy5 (1:320, Cat#301010, BioLegend, clone RPA-T8), anti-CD19-PE/Cy5 (1:160, Cat#302210, BioLegend, clone HIB19), anti-CD20-PE/Cy5 (1:160, Cat#302308, BioLegend, clone 2H7), anti-CD235ab-PE/Cy5 (1:320, Cat#306606, BioLegend, clone HIR2).

For GTAC, the only difference was that anti-CD38-BV711 (1:20, Cat#303526, BioLegend, clone HIT2) was used instead of anti-CD38-BV421.

Cells were stained for 30-45 min on ice, washed with FACS buffer, centrifuged, and resuspended in FACS buffer containing either 1:10^2 7-AAD (Cat#420404, BioLegend; for TARGET-seq+) or 1:10^5 DAPI solution (Cat#62248, Thermo Fisher Scientific; for GTAC).

2.6 Bulk cell sorting of primary human samples

FACS analysis and sorting for all experiments were performed on a MA900 Multi-Application Cell Sorter (Sony).

For TARGET-seq+ (Chapter 3), samples from different time points were processed in the same sort (except for pt5, where the end of cycle 11 (EOC11) sample was processed and sorted separately). Unstained cells, single-stained cells or beads, and cells stained with fluorescence-minus-one (FMO) mixes, were used as FACS controls to adjust the compensation matrix. For each sort, FACS sorting gates were defined using a combination of FMO controls and a stained healthy WT BM sample, in which BM population sizes were known *a priori*. For samples with a considerable amount of Lin^+ cells (usually > 70%), live/Lin^-CD34^+, live/$Lin^-CD34^-CD117^+$, and live/$Lin^-CD34^-CD117^-$ populations were first pre-sorted into 1.5 mL tubes containing FACS buffer with 1:10^2 7-AAD, on 'Yield' sorting mode. These cells were then subject to single-cell index sorting into 384-well plates (Cat#4ti-0384/C, Azenta Life Sciences), as described in the section 2.7.2. For GTAC experiments on primary human AML (Chapter 5), ≈85% of $CD34^+$ and ≈15% of Lin^+ cells were pre-sorted on 'Purity' sorting mode into 1.5 mL tubes IMDM with 10% FBS. The rest of the procedure was as described above.

In parallel, flow cytometry data was collected for each sample, prior to enrichment. Flow cytometry analysis was performed using FlowJo v10.8.1 and R version 4.3.1.

2.7 TARGET-seq+

2.7.1 TARGET-seq+: primer design and validation

Targeted genotyping primers for TARGET-seq+ (IDT, custom design), including primers used in RT-PCR (pre-amplification primers; Table A.3) and primers used in

genotyping PCR1 (nested to pre-amplification primers; Table A.4), were designed as described previously [80] (Appendix A.2), but using Primer BLAST [339]. Briefly, pre-amplification primers were designed to generate amplicons of 200–800 bp. When possible, two locus-specific primer pairs were designed: one amplifying the region from gDNA, and the other amplifying the region from cDNA. Where possible, PCR1 primers were designed to be nested to pre-amplification primers. For each PCR1 primer pair, the start of at least one primer was within 140 bp from the mutant nucleotide. All primers were first tested in bulk PCRs, followed by single-cell testing in a simulation of the TARGET-seq+ protocol (below) on primary human $CD34^+$ cells. Primer pairs which generated strong and specific genotyping bands whilst not compromising cDNA library generation were used in TARGET-seq+ experiments.

2.7.2 TARGET-seq+: lysis buffer preparation, single-cell sorting, and RT-PCR

TARGET-seq+ procedures were performed as described previously [80](Appendix A.2). The Mantis liquid dispenser (Formulatrix) with a high-volume chip (Cat# MCHVSMR6, Formulatrix) and the Viaflo electronic pipette (INTEGRA Biosciences) were used for liquid dispensing throughout the protocol.

Briefly, barcoded lysis buffer was prepared containing 0.5 mM dNTPs (Cat#19155, Life Technologies), 0.1% Triton X-100 (Cat#T8787, Sigma), 0.5 U/μL RNase inhibitor (Cat#2313B, Takara), 5% polyethylene glycol 8000 (PEG 8000; Cat#P1458, Sigma), 2.7 x 10^{-05} AU/μL protease (Cat#19155, Qiagen), 1:8,000,000 ERCC RNA spike-in mix (Cat#4456740, Invitrogen), and 668 nM of barcoded oligodT-ISPCR primer (sequences from [329]). 3 μL of barcoded lysis buffer were transferred into each well of 384-well plates. 384-well plates containing barcoded lysis buffer were frozen and kept at –80°C.

The next day, primary human samples were stained with FACS antibodies and pre-sorted as detailed above. Single, live cells were then index sorted into 384-well plates containing barcoded lysis buffer. Specifically, live/Lin^-CD34^+ haematopoietic stem and progenitor cells (HSPCs), live/$Lin^-CD34^-CD117^+$ GM precursor cells, and live/$Lin^-CD34^-CD117^-$ mature myeloid cells were sorted.

Additionally, for pt4, pt9, pt11, pt14, and pt15, HSPC sub-populations were enriched. In addition to the sample of interest, BM cells from the NOC156 control sample were sorted into ≈10% of every plate. Prior to sorting, the stream was first carefully calibrated for single-cell sorting, as detailed previously [330, 333] (Appendix A.4). The sorting procedure for each plate was carried out in 15-20 min, to avoid RNA degradation. After sorting, each plate was centrifuged, snap frozen on dry ice, and stored at –80°C. The next day, plates were incubated at 72°C for 15 min, followed by addition of 1 μL of the reverse transcription (RT) mix, bringing the reaction to 2.5 mM $MgCl_2$ (Cat#AM9530G, Invitrogen), 30 mM NaCl (Cat#AM9760G, Invitrogen), 25 mM Tris-HCl pH 8.0 (Cat#15893661, Thermo Fisher Scientific), 8 mM Dithiothreitol (DTT, Cat#707265ML, Thermo Fisher Scientific), 1 mM GTP (Cat#R1461, Thermo Fisher Scientific), 2 μM of Smart-seq2 template switching oligo (TSO, IDT, RNase-free HPLC purification; sequence: AAGCAGTGGTATCAACGCAGAGTACATrGrG+G), 0.5 U/μl RNase inhibitor, 2 U/μl of Maxima H-minus reverse transcriptase (Cat#EP0753, Thermo Fisher Scientific), and, where applicable, 70 nM of target-specific cDNA primers (IDT, RNase-free HPLC purification). The RT program was as follows: 42°C for 90 min, 10 cycles of 50°C for 2 min and 42°C for 2 min, and a final incubation at 85°C for 5 min. Following RT, 6 μL of PCR mix were added to achieve 1 × KAPA HiFi HotStart Ready Mix (Cat#07958935001, Roche), 50 nM ISPCR primer (IDT, RNase-free HPLC purification; sequence: AAGCAGTGGTATCAACGCAGAGT), 400 nM genotyping gDNA primers (IDT, RNase-free HPLC purification), and, where applicable, 28 nM genotyping cDNA primers. The PCR program was as follows: 98°C for 3 min, 21 cycles of 98°C for 20 s, 67°C for 30 s, and 72°C for 6 min, followed by incubation at 72°C for 5 min. The resulting cDNA/genotyping library was split into two aliquots: one was used for 3' biased whole transcriptome library construction while the other was used for single-cell genotyping library construction.

2.7.3 TARGET-seq+: 3' biased whole transcriptome library construction

3' biased whole transcriptome library preparation was performed as described previously [80] (Appendix A.2). This strategy preserves the cell barcode introduced by the oligodT at the 3' end whilst allowing to attach plate barcodes at the 5' end of the fragment. Briefly, 1 μL of cDNA/genotyping library was pooled from each well, using a Mosquito HTS liquid handling platform (TTP Labtech), generating plate-specific cDNA pools. Each cDNA pool was purified twice using Agencourt AMPure XP beads (Cat#A63881, Beckman Coulter), quantified by Qubit dsDNA HS Assay (Cat#Q32854, Life Technologies), and checked with either a High Sensitivity DNA Kit on a Bioanalyzer (Cat#5067-4626, Agilent) or a High Sensitivity NGS Fragment Analysis Kit (1 bp - 6,000 bp; Cat#DNF-474-0500, Agilent) on a Fragment Analyzer 5200. The Nextera XT DNA Library Preparation Kit (Cat#FC-131-1096, Illumina) was adopted for tagmentation-based library preparation, using a custom strategy to generate 3' biased libraries which retain oligodT-linked well barcodes [80]. 4 ng from each pool (5 μL) was mixed to 5 μL of ATM enzyme and 10 μL of TD buffer. The reaction was incubated at 55°C for 10 min and stopped with 5 μL of neutralization buffer. 5 μL of 2 μM i5 indexed primers (IDT; generic sequence: AATGATACGGCGACCACCGAGATCTACAC[barcode]GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTTGCT; barcodes used: CTCTCTAT; TATCCTCT; AGAGTAGA; GTAAGGAG; AAGGAGTA; CTAAGCCT; TCGACTAG; GCGTAAGA), 5 μL Nextera XT i7 forward index primer (Cat#FC-131-2001 and Cat#FC-131-2003, Illumina), and 15 μL of NPM enzyme were added, and the following PCR program was used: 95°C for 30 s, 14 cycles of 95°C for 10 s, 55°C for 30 s and 72°C for 30 s, and a final incubation at 72°C for 5 min. Libraries were purified twice with AMPure XP beads (0.7:1 volume ratio of beads to sample), quantified with Qubit, and checked for appropriate fragment size distribution on a Bioanalyzer. Libraries were pooled at equimolar ratios, the pool was diluted to 4 nM, and stored prior to sequencing. Libraries displaying TSO concatemers were excluded. Custom sequencing primers, diluted to

300 nM in HT1 buffer, were used for Read1 (IDT; PAGE purification; sequence: GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTTGC*T) and for Index2 (IDT; PAGE purification; sequence: AGCAACTCTGCGTTGATACCACTGCTTC-CGCGGACAGG*C). Libraries from all patients except pt5 were sequenced on a NovaSeq S4 flow cell (Illumina) with a depth of ≈1 million reads/cell, using the following configuration: 15 bp R1; 8 bp index read 1; 8 bp index read 2; 200 bp R2. Libraries from pt5 were sequenced on a NextSeq Mid Output v2.5 Flow Cell (Cat#20024905, Illumina) with a targeted sequencing depth of ≈40,000 reads/cell and the following configuration: 15 bp R1; 8 bp index read 1; 8 bp index read 2; 69 bp R2.

2.7.4 TARGET-seq+: single-cell genotyping library construction

Single-cell genotyping library construction was carried out as detailed previously [80] (Appendix A.2).

Briefly, two PCR steps (PCR1 and PCR2) generate fully indexed single-cell genotyping libraries for sequencing, starting from pre-amplified genotyping fragments generated during RT-PCR. PCR1 primers with CS1 (forward primer) and CS2 (reverse primer) adapters amplify initial genotyping fragments and insert a 6 bp plate barcode, distinguishing amplicons from different plates. In PCR2, sequencing adapters, containing a 10 bp well index, are attached to the PCR1 product, by annealing to the CS1/CS2 adapters.

A 6.5 μL PCR1 reaction was set up using 1.5 μL of cDNA/genotyping library, 3.25 μL of KAPA 2G Robust HS Ready Mix (Cat#KK5702, Sigma), and 300 nM of barcoded PCR1 primers (IDT, standard desalting). The following program was used: 95°C for 3 min, 20 cycles of 95°C for 15 s, 60°C for 20 s and 72°C for 1 min, and a final incubation at 72°C for 5 min. A 6.2 μL PCR2 reaction was set up using FastStart High Fidelity polymerase (Cat#4738292001, Sigma), 1.2 μL of well-indexed primer (Access Array Barcode Library for Illumina Sequencers- 384, Single Direction, Cat#100-4876, Fluidigm), and 1 μL of PCR1 product. The following program was used: 95°C for 10 min, 10 cycles of 95°C for 15 s, 60°C for 30 s and

72°C for 1 min, and a final incubation at 72°C for 5 min. 1 μL from each well was pooled using the Mosquito, keeping the plates separate. Libraries were purified with AMPure XP beads (0.8:1 volume ratio of beads to sample), quantified with Qubit, and checked using a Tapestation HS D1000 kit (Cat#5067-5583 and Cat#5067-5584, Agilent). Libraries were pooled at equimolar ratios, the pool was then diluted to 4 nM, and stored prior to sequencing. Libraries were sequenced on a NextSeq 500/550 Mid Output v2.5 kit (300 cycles), aiming for a depth of 2,000 reads per cell per amplicon, using 150 bp paired-end reads, with 10 bp for the cell barcode index read and custom sequencing primers (IDT, HPLC purification; sequences from [329].

2.8 TARGET-seq+ analysis

2.8.1 TARGET-seq+: single-cell genotyping data pre-processing

Single-cell genotyping reads were pre-processed as described previously [80] (Appendix A.2). Briefly, reads were demultiplexed via plate and well barcodes. For amplicons containing SNVs, alignment to human genome build 38 (hg38) was performed with STAR version 2.7.3a (default settings). Where applicable, cDNA and gDNA amplicons were separated into different BAM files, thus producing cDNA- and gDNA-specific readouts. Variant calling was performed with *mpileup* (samtools version 1.1, options: –minBQ 30, –count-orphans, –ignore overlaps). For amplicons containing small indels, WT and indel reads were counted from single-cell FASTQ files using the *grep* function in Unix. Coverage was computed as sum of WT and indel reads. In the case of *FLT3*-ITD, the getITD pipeline [340] was used, which was adapted for single-cell analysis by Dr. Bilyana Stoilova.

2.8.2 TARGET-seq+: mutation calling

Single-cell mutational calling was performed separately for each amplicon, as described previously [80] (Appendix A.2). Briefly, single-cell coverage was computed by summing all reads mapped to the mutant locus. Cells with coverage < 30 reads were excluded (i.e. the amplicon is called undetected). Rarely, this threshold was relaxed due to low coverage across most cells. Specifically, this applies to

the *NPM1* gDNA amplicon in pt11 and pt14 (threshold: 15 reads), *DNMT3A* gDNA/cDNA amplicons in pt11 (threshold: 5 reads), the *TP53* cDNA amplicon in pt15 (threshold: 10 reads), and the *BRINP3* gDNA amplicon in pt10 (threshold: 15 reads). Single-cell variant allele frequency (scVAF) for each cell was calculated as:

$$scVAF = mutant\ reads\ /\ coverage$$

Due to uneven amplification of the two alleles within single cells, a distribution of scVAFs is observed in heterozygous mutant cells. Furthermore, during library construction and sequencing, errors can be introduced which need to be accounted for when calling mutations. To define scVAF thresholds for calling a cell mutant, a previously published strategy [80, 332] (Appendix A.2) was adopted. Briefly, single cells from a control BM sample (WT at all genotyped loci) were sorted onto every 384-well plate and genotyped simultaneously with patient cells. The mean scVAF within these control cells (m) and the standard deviation (SD_m) were used to set genotyping thresholds at each locus. The scVAF threshold below which a cell was called WT is:

$$WT\ threshold = m + 3 \times SD_m$$

The scVAF threshold above which a cell was called mutant is:

$$MUT\ threshold = WT\ threshold + 0.01$$

If the scVAF was above this threshold and if the cell yielded ≥ 10 mutant reads, the cell was called mutant (the minimum mutant read number was relaxed for the few cases with poor coverage noted above). Cells with a scVAF between the two thresholds or where the number of mutant reads was < 10 were called undetermined.

If cDNA reads were available for a locus, the above strategy was performed separately for gDNA and cDNA amplicons. If the mutation was detected from

either gDNA or cDNA, the cell was called mutant. If the gDNA amplicon was WT and the cDNA amplicon was not mutant, the cell was called WT. If the gDNA amplicon failed single-cell genotyping and the cDNA amplicon was WT, the cell was called undetermined.

2.8.3 TARGET-seq+: FACS index data

During single-cell index sorting, fluorescence signal was recorded for forward scatter (FSC), back-scatter (BSC), Lineage cocktail/viability stain, and for all surface markers profiled via antibody staining, for each sorted cell. Cells were assigned as positive or negative for each surface marker, based on FACS gating. An immunophenotypic population was then assigned for each cell, based on the combination of readouts for each surface marker. As an example, $Lin^{-}CD34^{+}CD38^{-}CD10^{-}CD45RA^{-}CD90^{-}$ cells were assigned as immunophenotypic MPP. Immunophenotype for each cell was merged to its genotype, using matched plate and well indexes introduced via single-cell genotyping.

2.8.4 TARGET-seq+: transcriptome data pre-processing and quality filtering

Transcriptome sequencing data for pt4, pt5, pt9, pt14, and pt15 were pre-processed as described previously [80] (Appendix A.2). The pipeline was originally written by Dr. Niels Asger Jakobsen. Briefly, data was demultiplexed using plate and well indexes with bcl2fastq (v2.20) and cutadapt (v3.4)[341]. cDNA reads were simultaneously trimmed for Nextera adapters, poor-quality reads, and polyA tails. Reads were mapped to hg38 and ERCC92 transcripts with STARsolo (v2.7.10a)[342], and the GENCODE v38 reference gene annotation [343], which included only protein-coding genes and genes for long non-coding RNAs, was used. Gene counts were computed with default parameters and –soloType SmartSeq –soloFeatures GeneFull_Ex50pAS'. Metrics were computed with Samtools flagstat (v1.12), FastQC (v0.11.9), MultiQC (v1.11), and STAR outputs. Subsequent quality filtering was performed using Seurat (v5.0.1)[344] and SingleCellExperiment (v1.24.0)[345]. Cell-specific metadata, containing genotype and immunophenotype information, was

first merged with scRNA-seq data using single-cell barcodes. Cells with $\geq$ 25,000 reads assigned to genes, $\geq$ 2,000 detected genes, $\leq$ 50% reads assigned to ERCC RNA, and $\leq$ 15% mitochondrial reads, were used to generate a Seurat object with raw counts. This analysis was performed in collaboration with Felix A. Radtke.

2.8.5 TARGET-seq+: copy number alteration inference

In pt4, pt5, pt9, pt14, and pt15, where CNAs were identified by karyotyping, CNAs were inferred in single cells from RNA expression, using inferCNV (v1.18.1) [322, 346]. For each patient, *CreateInfercnvObject* was applied, providing the patient-specific counts matrix containing all the cells from the respective patient and all healthy control BM cells passing scRNA-seq quality filters. Next, the *infercnv* command was ran with the following parameters: cutoff = 1, denoise = TRUE, HMM = TRUE, analysis_mode = "subclusters", hclust_method = "ward.D2", leiden_resolution = 0.01, tumor_subcluster_pval = 0.05. To identify high-confidence CNAs, only those CNAs detected by bulk karyotyping (Table 3.3) were interrogated (with the exception of chromosome 7 loss in pt4, which showed a strong CNA signal but was not detected by karyotyping). The resulting set of regions containing *bona fide* CNAs was subjected to single-cell analysis. Specifically, for each single cell, it was calculated whether the average gene expression in the region of interest (either the entire chromosome for large CNAs, or a manually selected portion of the chromosome for shorter CNAs) was significantly higher or lower than expected in a diploid cell. The standard deviation of the average read count per cell and per gene, across the entire genome (which, on average, is diploid in the analysed samples), was calculated as

$$\sigma_{<n>} = \sqrt{G} \times \sqrt{\frac{1}{k} \times \sum_{k=1}^{K} (\langle \mathrm{n} \rangle_{(G,k)} - \mu_{<n>G})^2}$$

where G is the total number of genes used for analysis, K is the total number of cells (both taken from the final infercnv object), $\langle \mathrm{n} \rangle_{(G,k)}$ is the average number of normalized read counts per gene in cell k, and $\mu_{<n>G}$ is the average of $\langle \mathrm{n} \rangle_G$ across

all cells (both estimated from all G genes). Single cells were classified as affected by a particular CNA if gene expression in the respective region was significantly higher (copy number gains) or lower (copy number losses) than expected by chance. To this end, the average gene expression per cell was computed on the subset of genes, g, affected by a particular CNA of interest (specified in the respective figure captions), yielding an estimate $\langle \mathrm{n} \rangle_{(g,k)}$ for each cell k. Accepting a false discovery rate (FDR) of 5% and assuming that average read counts per cell are normally distributed, cells were then classified as follows:

$$\text{harbouring a copy number gain if:} \langle \mathrm{n} \rangle_{(g,k)} > \mu_{<n>G} + 1.96 \times \sigma_{<n>} \frac{}{\sqrt{g}}$$

$$\text{harbouring a copy number loss if: } \langle \mathrm{n} \rangle_{(g,k)} < \mu_{<n>G} - 1.96 \times \sigma_{<n>} \frac{}{\sqrt{g}}$$

and else as diploid. Single-cell CNA calls were then merged with SNV/small indel calls for the same cells, and clonal hierarchies were established as detailed in section 2.8.6.

Of note, transcriptome libraries for pt5 were sequenced at lower depth compared to other patients and control cells (counts on average 14,369 versus 1,919,130, respectively). To enable accurate CNA calling for pt5, probabilistic downsampling of the healthy control BM cells was performed. A scaling factor was calculated by dividing the average number of counts across cells from pt5 and healthy control BM cells sequenced with pt5 by the average number of counts across cells from healthy control BM samples sequenced to a higher depth. A probability distribution was then calculated by dividing the counts for a given gene by the total number of counts. These probabilities were used to weight the random selection of counts to maintain an accurate representation of count distributions across genes. Downsampling was performed in collaboration with Felix A. Radtke.

Downsampling enabled an efficient comparison between cells from pt5 and those from a control BM originally sequenced at higher depth. CNA calling in pt5 was performed as detailed above but accepting an FDR of 10%, given the weaker signal owing to a low sequencing depth. Genome-wide inferCNV heatmaps are shown in Figures A.1, A.2, A.4, A.3, A.5.

2.8.6 TARGET-seq+: inference of clonal hierarchies

Mutation calls across all screened loci and, where applicable, CNA calls, were first integrated for each single cell in each patient. The pattern of somatic event co-occurrence informed the most likely course of somatic events (clonal hierarchy) in each patient, as previously described [80, 270, 332]. Briefly, infSCITE [347] was used with default parameters and '-r 200 -L 10000 -fd 0.01 -ad 0.02 -e 0.2 -p 1000', to obtain the statistically most probable phylogenetic tree. Based on the infSCITE output, each cell was assigned to a clone.

In two patients (pt10 and pt11), the exact phylogenetic tree could not be confidently assigned. In pt11, a substantial number of cells failed single-cell genotyping for the *DNMT3A*-N757K locus and infSCITE identified the *DNMT3A*-N757K mutation as subclonal to the *NPM1*-W288fs*12 mutation. However, this is unlikely, due to *DNMT3A* mutations most frequently occurring in the founder clone [280]. Bulk WES corroborated both mutations being early events (VAF of 47% for *NPM1*-W288fs*12 and 44% for *DNMT3A*-N757K). In pt10, the infSCITE output was unreliable (predicted "ad" value of 0.41). In neither case, the resulting uncertainty impacted the conclusions.

In two cases, further manual curation of the phylogenetic trees was performed following infSCITE analysis. In pt15, infSCITE indicated the two *NF1* mutations as subclonal to *IDH1*-R132C. However, longitudinal NGS did not detect the *IDH1*-R132C at relapse, while the *NF1* mutations were both detected, indicating *NF1* mutations were harboured by *IDH1*-WT clones. In pt4, the inferCNV data, combined with single-cell genotyping data, suggested that gain of chromosome 1 (chr1+) occurred twice, in two independent clones: once in the SIA(-20q) clone, and once in the SIAJ clone. Due to chr1+ being treated as a single event in the infSCITE input matrix, the two clones were assigned manually.

ADO is a technical artefact in single-cell genotyping, in which one allele fails PCR amplification. If ADO occurs on a mutant allele in a heterozygous mutant cell, the cell is wrongly called WT. If ADO occurred in an early mutation, but a later 'control' mutation was detected, the cell was still assigned to the appropriate

daughter clone (inferring ADO of the ancestral mutation occurred). For example, in pt18, there were 14 cells mutant for *SRSF2* but WT for *IDH1*; given that the *IDH1* mutation occurred prior to the *SRSF2* mutation (as predicted by infSCITE), these cells were still assigned to the *SRSF2*/*IDH1*-mutant clone. In cases where a branching clonal structure was present, a small proportion of cells occasionally contained mutations which were mutually exclusive as per clonal structure. These cells were excluded.

2.8.7 TARGET-seq+: inference of clone size in Lin$^-$ cells

For pt4, pt5, pt9, pt10, pt11, pt14, and pt18, the size of each clone within all live, Lin$^-$ cells at each time point was inferred. For each sample, the fraction of each clone within each immunophenotypic population was computed, and the upper and lower limits of the 95% confidence interval (CI) were calculated via the Wilson score interval. Next, both the population-specific clonal fractions and the upper and lower limits of the CI were multiplied with the relative size of the respective immunophenotypic population (measured by flow cytometry as the fraction of that population within Lin$^-$ cells). Population-specific clonal fractions were summed for each clone, to infer the total clonal size within Lin$^-$ cells. The CI limits (multiplied with the relative size of the respective immunophenotypic population) were also summed for each clone, to compute the 95% CI of the inferred clone size within Lin$^-$ cells. The results for each patient are summarized in Tables A.6 - A.12. In all cases in which one or several Lin$^-$ were not sampled, those populations contributed cumulatively to $< 1\%$ of all Lin$^-$ cells. This analysis was not done for pt15 and pt20, in which < 5 cells were sampled from the largest Lin$^-$ population at a specific time point, precluding high-confidence analysis.

2.9 Genotyping integration to bulk ATAC

To validate whether targeted genotyping is implementable into the ATAC chemistry, two strategies were employed: one using standard bulk Omni-ATAC

[348] and the other simulating the single-cell ATAC-seq (scATAC-seq) strategy [349, 350] in bulk.

For the first strategy, ≈50,000 OCI-AML3 cells were harvested and washed twice with 1 mL of ice-cold PBS (Cat#14190169, ThermoFisher), centrifuging at 500 × g, 4°C, for 5 min, in a pre-chilled fixed angle centrifuge. Cells were resuspended in 50 μL of ATAC resuspension-detergent buffer (ATAC-RSB-DTN) buffer, containing 0.01% digitonin (Cat#G9441, Promega), 0.1% NP-40 (Cat#I8896, Sigma, for the IGEPAL CA-630 equivalent), 0.1% Tween-20 (Cat#655205-250ML, Sigma), 10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 3 mM $MgCl_2$, and incubated on ice for 3 min. Next, 1 mL of ATAC-RSB containing only 0.1% Tween-20 was added, tubes were inverted slowly 3 times, and nuclei were centrifuged at 500 × g at 4°C for 10 min in a fixed-angle centrifuge. The supernatant was aspirated carefully, not disrupting the nuclear pellet. 50 μL of tagmentation mix, containing 25 μL of 2X TD buffer (Cat#20034197, Illumina), 16.5 μL of PBS, 0.5 μL of 1% digitonin, 0.5 μL of 10% Tween-20, and 2.5 μL of Illumina Tn5 enzyme (Cat#20034197, Illumina), were added by slowly pipetting 6 times, and nuclei were tagmented for 30 min at 37°C with 1000 rpm shaking on a ThermoMixer C (Cat#15158953, Thermo Fisher Scientific). DNA cleanup was performed with the NEB Monarch PCR & DNA Cleanup Kit (5 μg)(Cat#T1030L, New England Biolabs). Transposed DNA was eluted in 20 μL of elution buffer. The ATAC PCR was set up by mixing 19 μL of transposed DNA, 1 μL of nuclease-free water, 2.5 μL of 25 μM ATAC forward primer (IDT; CAAGCAGAAGACGGCATACGA-GATTCGCCTTAGTCTCGTGGGCTCGGAGATGT), 2.5 μL of 25 μM ATAC reverse primer (IDT; AATGATACGGCGACCACCGAGATCTACACTCGTCG-GCAGCGTCAGATGTG), and 25 μL of NEBNext High-Fidelity 2X PCR Master Mix (Cat#M0541L, New England Biolabs). For the genotyping condition, 0.5 μL of 50 μM forward genotyping primer (IDT; CCCAGGGTATTTGGTTTCCCAG) and 0.5 μL of 50 μM reverse genotyping primer (IDT; ACTCCTTCAGCGGAGCGAAG) against *DNMT3A*-R882 were added instead of water. The following PCR program was used: 72°C for 5 min, 98°C for 30 s, then 18 cycles of 98°C 10 s, 63°C 30 s

and 72°C for 1 min. The PCR product was purified with the NEB Monarch PCR & DNA Cleanup Kit (5 μg). ATAC libraries were checked on Bioanalyzer. An aliquot of the product was used in a second PCR to check for genotyping target amplification. This reaction was set up from 10 μL of PCR product, 15 μL of NEBNext High-Fidelity 2x PCR Master Mix, 2.5 μL of 6 μM forward genotyping primer, and 2.5 μL of 6 μM reverse genotyping primer. The following PCR program was used: 98°C for 30 s, then 25 cycles of 98°C 10 s, 63°C 30 s and 72°C for 1 min. Amplicons were checked on a 2% agarose gel.

For the second strategy, the protocol was as above until the tagmentation reaction. After tagmentation, 50 μL of tagmentation stop buffer (20 mM of Ethylenediaminetetraacetic acid, (EDTA) pH 8.0 (Cat#AM9260G, Thermo Fisher Scientific) and 10 mM Tris-HCl pH 8.0) were added, the suspension mixed 10 times, and left on ice for 10 min. Nuclei were centrifuged at 500 $\times$ g at 4°C for 10 min in a fixed-angle centrifuge. The supernatant was removed, and 23.25 μL of lysis buffer were added, containing 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.036 AU/mL Protease, and 0.2% sodium dodecyl sulfate (SDS; Cat#V6551, Promega). Nuclei were incubated at 65°C for 15 min. Next, 23.25 μL of Tween-20 10% was added to neutralize SDS and nuclei were incubated at room temperature for 10 min. A PCR was set up by mixing 46.5 μL of sample, 50 μL of NEBNext High-Fidelity 2x PCR Master Mix, 1 μL of nuclease-free water, 1.25 μL of 100 μM ATAC forward primer, and 1.25 μL of 100 μM ATAC reverse primer. For the genotyping condition, 0.5 μL of 100 μM forward genotyping primer and 0.5 μL of 100 μM reverse genotyping primer were added instead of water. The rest of the protocol was performed as detailed above.

2.10 GTAC

2.10.1 GTAC: primer design and validation

The design of pre-amplification and nested genotyping primers used for GTAC is similar as in TARGET-seq+ (2.7.1), with few modifications [332, 333](Appendix A.4). First, only gDNA genotyping primers were used. Second, genotyping amplicons

were in most cases below 300 bp in length, to avoid Tn5 tagmentation within the amplicon. In rare instances (genotyping of the *NPM1* mutation), the amplicon was > 300 bp due to the inability to design efficient primers generating a shorter amplicon. Third, pre-amplification primers were always checked with the IDT Oligo Analyzer for annealing to i7/i5 ATAC primers, and those displaying $\Delta G < -9$ kcal/mol were excluded. Fourth, nested genotyping (PCR1) primers were always designed so that the beginning of at least one primer was within 69 bp from the mutation.

All primers were first tested in bulk PCRs simulating GTAC conditions. Primer pairs generating strong amplicons were tested in single cells, in a simulation of the GTAC protocol (below). For testing purposes, OCI-AML3, K562, or primary human $CD34^+$ BM cells were processed in 96-well plates and non-indexed lysis buffer. Testing volumes were doubled compared to standard procedures (below). Each primer pair was tested on at least 16 single cells. After 18 cycles of PCR, pooled scATAC-seq libraries were checked on a Bioanalyzer. Single-cell amplicons were subject to further 32 cycles of PCR and checked on a 2% agarose gel. Prior to GTAC tests on any cell type, scATAC-seq without genotyping was carried out first to determine the optimal number of PCR cycles via qPCR, using SYBR Green I (Cat#S7563, Thermo Fisher Scientific) [349].

2.10.2 GTAC: experimental procedure

GTAC was performed as detailed previously [332, 333](Appendix A.3 for the original protocol; Appendix A.4 for a detailed step-by-step procedure). Briefly, 1X lysis buffer was prepared containing 10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% SDS, 5 μM i7 well barcodes (IDT, [350]) and 5 μM i5 plate barcodes (IDT, [350]. 2 μL of 1X lysis buffer were added to each well of 384-well plates, which were frozen and stored at -80°C.

Within 1 month, primary BM human samples were thawed, stained, and pre-sorted, as described in the sections above. 50,000-90,000 cells were FACS-sorted into 1.5 mL Eppendorf tubes pre-coated with 0.5% bovine serum albumin (BSA;

Cat#A8577, Sigma) and subject to lysis and Tn5 tagmentation, as described previously [333]. Following tagmentation, 50 μL of tagmentation stop buffer (10 mM EDTA, 0.5% BSA) were added and nuclei left on ice for 10 min and a maximum of 4 h. Next, 500-1500 μL of nucleus dilution buffer (0.5% BSA, 0.5X PBS in nuclease-free water) were added. DAPI was added for a final concentration of 1:10^5. The MA900 multi-application cell sorter was calibrated for single nucleus sorting [333](Appendix A.4). Single nuclei were identified using a combination of FSC, SSC and DAPI fluorescence, and sorted into 384-well plates containing 1X lysis buffer. Sorted plates were frozen and stored at -80°C.

Within 1 week, plates were incubated at 65°C for 15 min. Next, 1 μL Tween-20 10% was added to each well. The PCR was set up by adding 5 μL of mix containing 2X NEB High Fidelity Q5 master-mix (Cat#M0515, New England Biolabs; alternatively, Q5 high-fidelity 2X master mix (Cat#M0492L, New England Biolabs)) and pre-amplification genotyping primers (IDT; final concentration 100 nM per primer). The following PCR program was used: 72°C for 10 min, 98°C for 1 min, then 16-18 cycles of 98°C 10 s, 60°C 30 s and 72°C for 30 s, and a final elongation at 72°C for 5 min. 18 cycles were used for cell lines and 16 cycles for primary human cells. Next, 20% of the material was transferred to new plates containing 4 μL of nuclease-free water (Cat#AM9932, Invitrogen), which were frozen and kept at -80°C until use. Hereafter, these plates are referred to as genotyping stock plates.

To purify scATAC-seq libraries, the remaining material was pooled from each well, keeping plates separate. The pools underwent column purification followed by double-sided AMPure XP bead cleanup, first with a 0.5:1, then with a 1.2:1 volume ratio of beads to sample (detailed in Appendix A.4). Pools were quantified with Qubit and scATAC-seq library trace checked with a Bioanalyzer or Agilent Tapestation (High Sensitivity D5000 ScreenTape and Reagents, Agilent, Cat#5067-5592 and Cat#5067-5593). Libraries were pooled at equimolar ratios, diluted to 4 nM, and stored until sequencing. K562 libraries (Chapter 4) were sequenced at ≈435,000 reads/cell. The AML sample was sequenced at ≈256,000 reads/cell.

Illumina NextSeq 500/550 Mid-Output or High-Output (Cat#20024907, Illumina) kits were used with 75 bp paired-end reads.

Single-cell genotyping PCR1 and PCR2 were performed starting from genotyping stock plates, similarly as in TARGET-seq+, with few modifications [332, 333]. First, PCR1 was performed using 22 cycles for cell lines and 24 cycles for primary human samples. Second, post-PCR2 genotyping pools were purified using AMPure XP beads with a 1.2:1 ratio of beads to product. Third, the final pool underwent an additional AMPure XP bead purification with a double-size selection, using first a 0.6:1, followed by a 1.2:1 ratio of beads to library. Libraries were sequenced at a depth of ≈2,000 reads per cell per amplicon. Illumina MiSeq kit v3 (Cat#MS-102-3001, Illumina) or NextSeq Mid Output kits (75 bp paired-end) were used. Three custom primers (IDT, HPLC purification) were used for genotyping library sequencing [333](Appendix A.4).

2.11 GTAC analysis

2.11.1 Single-cell genotyping analysis

Single-cell genotyping analysis for GTAC was performed similarly as for TARGET-seq+, with a few modifications [332](Appendix A.3). Pre-processing was carried out without cDNA/gDNA amplicon separation, given that genotyping was performed exclusively from gDNA. scVAF and coverage values were calculated as in TARGET-seq+. Cells with coverage < 30 reads were excluded. For cell line analysis, cell line(s) WT for the genotyped loci were used as controls for mutation calling. For mutational analysis of the human AML sample, 300 $CD34^+$ BM cells from an individual WT at all the genotyped loci were used as controls. The maximum scVAF within these control cells ($\max_{scVAF}$) and the SD of scVAF in WT control cells were used to set genotyping thresholds. The scVAF threshold above which a cell was called mutant is:

$$MUT\ threshold = max_{scVAF} + 2 \times SD$$

Given the maximum observed number of mutant reads in a single WT control cell ($\max_{mut}$), a threshold for the minimum number of mutant reads to call a cell mutant was calculated as:

$$Minimum\ number\ mutant\ reads = max_{mut} + 1$$

Cells were called mutant if both scVAF and mutant read number were above these thresholds. Clonal assignment and inference of the clonal hierarchy were performed as in TARGET-seq+. For the AML sample, if mutations were called on *TET2*, *ASXL1* and *RUNX1*, but both *BCOR* loci failed detection, these cells were removed. 18 cells in which both *BCOR* mutations were detected were also excluded.

2.11.2 GTAC: ADO estimation

ADO rates in GTAC were estimated for a range of loci. To this end, mutations/variants known to be present in a heterozygous state in all cells were analysed (i.e. there was a ground truth regarding their genotype). The frequency of homozygous calls (either reference or alternate) was computed to measure ADO rates. 5 heterozygous loci in K562 (*ASXL1*, *NOTCH1*, *CBL*, *HOXA9*, *TET2*), 2 in OCI-AML3 (*NPM1*, *DNMT3A*), 1 in JURKAT (*RUNX1*), and 1 in MOLM-13 (*FLT3*) cells were genotyped. In the primary AML sample, 2 heterozygous germline single-nucleotide polymorphisms (SNPs) in *SH2B3* and *ASXL2* were genotyped. WT cells were used to set scVAF thresholds (section 2.11.1). Given the maximum scVAF within these control cells ($\max_{scVAF}$) and the SD of scVAF in WT control cells, the scVAF threshold above which a cell was called mutant/variant is:

$$MUT/VAR\ threshold = max_{scVAF} + 2 \times SD$$

For cells with scVAFs below this threshold, I estimated ADO of the mutant/variant allele occurred. The mutant/variant ADO percentage at each locus was calculated as:

MUT/VAR ADO percentage = (n° cells below MUT/VAR threshold)/(total n° cells passing genotyping QC) × 100

To determine detection of the reference (or WT) allele, an inverse scVAF threshold was defined as:

inverse threshold = 1 - MUT/VAR threshold

For cells with scVAFs above this threshold, I estimated ADO of the reference allele occurred. The reference ADO percentage was calculated as:

REF ADO percentage = (n° cells above inverse threshold)/(total n° cells passing genotyping QC) × 100

The total ADO percentage was calculated as:

total ADO = MUT/VAR ADO percentage + REF ADO percentage

The percentage of cells showing detection of both reference and mutant/variant alleles was calculated as:

Percentage of cells with REF and MUT/VAR alleles detected = 100 - total ADO

2.11.3 GTAC: scATAC-seq data pre-processing

For detailed analytical procedures, please refer to [332](Appendix A.3). In this and subsequent sections, I provide a brief overview of the main analytical steps. Raw BCL files were demultiplexed using i7/i5 barcodes. The 3' Nextera sequence was removed with cutadapt [341]. Read alignment to hg38 was performed with HISAT2 [351]; reads showing mapping quality < 30 were removed. Non-deduplicated BAM

files were used to estimate coverage of genomic loci from scATAC-seq data using *bedtools coverage*. BAM files were de-duplicated using *MarkDuplicates* in Picard. Samtools was used for BAM file indexing and merging, and aggregated fragment size was computed on the merged file excluding mitochondrial reads. The merged BAM file was converted to BED format, used as input for MACS2 peak calling [352]. The resulting files were used as input for the UCSC genome browser [353], and for peak correlation analysis [354]. A count matrix was produced by counting the number of reads from individual cells overlapping the union peak set. Per-cell quality metrics were generated from individual BAM files, using MACS2 peak coordinates. Of note, whilst this pipeline calls peaks on all cells, for the analyses throughout Chapter 5, peaks were called within ArchR (v1.0.2) [355], using a custom strategy (sections below).

2.11.4 GTAC: quality filtering

ArchR was used for downstream GTAC analyses. Fragment files were used to generate Arrow Files. Cells with $>$ 120,000 and $<$ 10,000 unique nuclear fragments, and transcription start site (TSS) enrichment $<$ 8, as well as predicted doublets, were excluded. To merge genetic clonal identity with scATAC-seq data for each cell, genotyping data was integrated to scATAC-seq metadata using shared cell IDs. Specifically, each unique i5/i7 combination (scATAC-seq) matched a unique plate/well barcode combination (single-cell genotyping). Genetic clonal information for each cell was added to the ArchR metadata, as detailed previously [332]. Cells not assigned to a genetic clone were removed.

2.11.5 GTAC: dimensionality reduction, clustering, and cluster annotation

Latent Semantic Indexing (LSI) dimensionality reduction [356], cell clustering, and Uniform Manifold Approximation and Projection (UMAP) generation were performed as previously described [332, 355]. No batch effect correction was applied. Clusters were annotated via several strategies [332](Appendix A.3). Briefly, the most accessible genes in each cluster (cluster marker genes) were computed by

comparing one cluster at a time to all other cells. Cluster marker genes were used to identify differentiation stages. Gene accessibility was visualized on the UMAP embedding, using Markov Affinity-based Graph Imputation of Cells (MAGIC)[357] for signal smoothing. Cluster marker genes with $\log_2$FC ≥ 0.25 and FDR ≤ 0.05 were screened for enrichment of human haematopoietic transcriptional signatures [81, 89, 90, 358–360] using g:Profiler [361]. Per-cell deviation scores for TF binding motif accessibility were computed with chromVAR [362] and visualized on the UMAP embedding using MAGIC. Cluster-specific differentially accessible peaks of open chromatin were screened for enrichment in published bulk ATAC-seq signatures of sorted haematopoietic populations (provided by ArchR). Finally, the data was projected onto a healthy human haematopoiesis scATAC-seq dataset [88] to confirm cluster assignment. The LSC score was computed using the combined accessibility of genes shown to be transcriptionally upregulated in LSCs compared to downstream leukaemic cells [244], using the *addModuleScore* function in ArchR.

2.11.6 GTAC: peak calling, marker peak identification, and co-accessibility

Cell pseudo-bulk replicates were generated by aggregating single cells belonging to distinct genetic clones [332](Appendix A.3), with the exception of the TA clone, which was assigned to only 19 cells. Using 5 clone-specific pseudo-bulk replicates, peaks were re-called using MACS2 within ArchR. This resulted in 175,563 peaks of open chromatin called within 3,113 cells from the AML sample. Clone- and cluster-specific marker peaks were then computed. Unless otherwise specified in the Figure captions, marker peaks were defined as those displaying FDR ≤ 0.05 and $\log_2$FC $\geq$ 1. Marker peaks were used to estimate clone- and cluster-specific TF binding motif enrichment and bulk ATAC-seq signature enrichment. Peak co-accessibility scores were added across the entire peak set to estimate how frequently the accessibility of two peaks co-varies in the same direction [355].

2.11.7 GTAC: differential accessibility analysis

Clone-specific analysis of chromatin accessibility was carried out using several approaches, detailed previously [332](Appendix A.3). Briefly, for pairwise comparisons of gene activity between clones, the $\log_2$FC was calculated using the Gene Score Matrix, and the FDR was computed via the Wilcoxon rank sum test with multiple testing correction. Pathway enrichment was performed following pairwise comparisons, by using fgsea (v1.20.0) [363] and the Hallmark, CGP, TFT, and GO MSigDB collections [364]. Similarly, pairwise comparisons of peak accessibility was carried out using the Peak Matrix. For each differentially accessible peak, the closest gene was computed by ArchR. To understand whether differentially accessible peaks were enriched in specific cis-regulatory elements, peaks were filtered for those mapping to either promoter-proximal regions (2 kb upstream to 100 bp downstream of annotated TSS) or to distal elements (all peaks mapping neither to TSS-proximal regions, exons, nor introns). For pairwise comparisons of TF binding motif accessibility, per-cell TF binding motif deviation scores (z-scores) were first computed using chromVAR. Z-scores are a measure of how much TF binding motif accessibility in a single cell deviates from the average of all cells, taking cell-specific sequencing depth into account. Differences in z-scores were computed in pairwise comparisons between clones, using the Wilcoxon rank sum test with multiple testing correction.

2.11.8 GTAC: projection on reference

Projection of the AML dataset onto a healthy human scATAC-seq haematopoiesis reference was performed as described previously [332](Appendix A.3). Briefly, the AML fragment file was converted to the hg19 annotation (used in the reference dataset). Peak calling was performed as described above. The 'unknown' clusters were removed from the reference, and peaks from the AML and reference datasets were intersected using Signac (v1.9.0) [365]. Peaks in the AML dataset were replaced with peak counts corresponding to the intersecting regions and LSI was re-computed on the new peak set in Signac. The anchor set on shared peaks was computed

between the AML and reference datasets using Seurat (v4.3.0), resulting in 3265 anchors. Reference cell labels were transferred to the AML dataset. The AML cells were then projected onto the reference UMAP embedding using the above anchors (the reference embedding was loaded with uwot (v0.1.14)). This analysis was performed in collaboration with Simone G. Riva.

2.11.9 GTAC: trajectory analysis

Differentiation trajectories were computed with a supervised approach within ArchR as described previously [332](Appendix A.3). Briefly, based on prior knowledge of haematopoietic differentiation, three potential differentiation pathways were identified within the leukaemic continuum. The starting point for each trajectory was the LMPP-like cluster, given its highest LSC score. Pseudotime vectors of cells across the three trajectories were computed with the *addTrajectory* function, assigning a pseudotime score to each cell. The *geom_density* function from ggplot2 (v3.4.0) quantified genetic clones across pseudotime. The *geom_smooth* function with loess method quantified clone-specific gene accessibility across pseudotime.

To analyse whether distinct clones displayed variable differentiation dynamics, the three differentiation pathways were also computed retaining only cells from a specific genetic clone. With this strategy, TF binding motif accessibility dynamics and chromatin variation were studied across a given clone-specific trajectory.

2.11.10 GTAC: identification of putative PRC1.1 direct targets

A list of genes frequently bound by KDM2B (the DNA-binding component of PRC1.1 [366]) in primary human AML was manually curated from a previously published chromatin immunoprecipitation with sequencing (ChIP-seq) dataset [367]. Putative PRC1.1 target genes in human AML were defined as those bound by KDM2B in at least 5/6 screened patients.

2.11.11 GTAC: TF footprinting analysis

Clone-specific TF footprinting was performed on clone pseudo-bulk replicates [332]. TF binding motifs footprints were computed with *getFootprints*, and *plotFootprints* was used to generate the graphs. To correct for an inherent binding preference of the Tn5, the Tn5 bias was subtracted to normalize each footprint. The Tn5 binding bias was graphically represented below the footprints.

Chapter 3

Clonal basis of resistance or response to ivosidenib combination therapies

Contents

3.1 Aims of Chapter 3 63
3.2 Patient cohort studied 64
3.2.1 Clinical characteristics, treatment course, and responses 64
3.2.2 Longitudinal immunophenotypic analysis of the patient cohort 66
3.2.3 Application of TARGET-seq+ to sequential patient samples 70
3.3 Newly detected genetic clones are selected within LSCs months or years prior to treatment failure 75
3.3.1 *FLT3*-ITD or *NF1*-mutant clones emerge within LSCs 3 cycles into treatment and associate with rapid relapse . 75
3.3.2 Resistant clones emerge years prior to treatment failure and expand slowly in a competing clonal milieu 79
3.4 Genetic clones detected at baseline undergo rapid therapy adaptation 82
3.4.1 A pre-exiting clone harbouring trisomy 4 loses differentiation capacity rapidly upon treatment exposure 83
3.4.2 A *FLT3*-mutated clone rapidly loses differentiation capacity upon treatment exposure 85
3.4.3 Pre-existing clones gradually lose differentiation capacity during treatment in a complex clonal milieu 87
3.5 Clonal foundations of sustained clinical response . . . 89
3.6 Summary of Chapter 3 and proposed model 93

I am grateful to Dr. Bilyana Stoilova and Angus Groom, who helped me with TARGET-seq+ experiments; to Marlen Metzner, who performed the 97-gene panel; and to colleagues from the MD Anderson Cancer Center, who performed the 81-gene panel and karyotyping.

3.1 Aims of Chapter 3

The current standard of care for *IDH1*-mutated AML patients is the combination of the mutant IDH1 inhibitor ivosidenib and the hypomethylating agent azacitidine [172, 192]. Interestingly, the combination of the BCL-2 inhibitor venetoclax with azacitidine is particularly efficient in *IDH*-mutated AML patients [168, 169, 197]. Hence, to identify more effective treatment regimens for *IDH1*-mutated patients, a recent phase Ib/II clinical trial evaluated the safety and efficacy of ivosidenib with venetoclax ± azacitidine in 31 *IDH1*-mutated patients with myeloid malignancies [199] (NCT03471260). The initial overall response rate (ORR) was 94% across all subgroups. After a median follow-up of 40.3 months (data cut-off 12/14/2023), 8 patients remained on study, 12 patients received allogeneic transplant, but 11 patients relapsed. Understanding the main drivers of relapse thus remains a major clinical need, and a detailed characterization of the genetic and cellular landscapes of resistant disease is required. Whilst some genetic correlates of relapse following ivosidenib and venetoclax ± azacitidine were previously identified [199], the clonal trajectories establishing resistance remain poorly understood. Specifically:

1. The kinetics of clonal trajectories leading to either resistance or sustained response, to identify the earliest post-therapy time points when treatment outcome may be determined, were not studied.

2. It is unclear in which BM compartments resistant clones are selected, whether selection is operative specifically within LSCs, and whether this is evident prior to clinically frank relapse.

3. The biological basis of treatment failure in patients relapsing without novel driver mutations remains poorly understood.

4. It is unclear why some patients relapse after months, whilst others relapse after years of treatment.

5. Finally, why treatment fails in some but not all patients, and what is the clonal basis of sustained response, remains unresolved.

To address these gaps, sequential BM samples from 9 patients were analysed with TARGET-seq+ [80], which simultaneously provides high rates of genotyping coupled with immunophenotype and RNA-seq at single-cell resolution. 7 of these patients relapsed, while 2 remained in sustained remission for 48 and 57 months. By analysing paired samples from baseline, early treatment response, and either from relapse or long-term remission, I show that the clonal milieu of immature cells early during treatment, while patients are in clinical remission, foreshadows patient outcome.

3.2 Patient cohort studied

3.2.1 Clinical characteristics, treatment course, and responses

9 patients (pts) (5 females and 4 males; Figure 3.1 and Table 3.1) had sufficient samples from each time point to be studied. 7/9 patients experienced treatment failure whilst 2/9 remained in sustained remission as of last data cut-off. 7 patients (pts 5/9/10/11/15/18/20) had AML at trial entry: 3 *de novo* AML; 1 secondary AML (sAML); and 3 relapsed/refractory AML (R/R AML). 2 patients (pts 4/14) had myelodysplastic syndrome/myeloproliferative neoplasm overlap syndrome (MDS/MPN). 6 patients (pts 4/5/9/10/11/14) received the ivosidenib and venetoclax doublet (2 received 400 mg of venetoclax and 4 received 800 mg of venetoclax on days 1-14; all received 500 mg of ivosidenib, continuously, initiated on cycle 1, day 15). 3 patients (pts 15/18/20) received ivosidenib, venetoclax, and azacitidine (400 mg of venetoclax on days 1-14, 500 mg of ivosidenib, continuously, initiated on cycle 1, day 15, and 75 mg/m2 azacitidine for 7 days). Pt20 received ivosidenib, venetoclax, and azacitidine for two cycles, followed by ivosidenib monotherapy, after prolonged cytopenia in the first two cycles. The median *IDH1* mutation VAF in the baseline BM MNCs was 22% [199](NGS targeted panel; Table 3.1). The treatment course and sample collection time points are shown in Figure 3.1.

Characteristic	pt11	pt15	pt20	pt5	pt9	pt14	pt4	pt18	pt10
Age at time of enrolment (years)	65	65	63	84	77	80	68	75	72
Sex	Female	Female	Male	Female	Male	Female	Male	Male	Female
Diagnosis	AML	R/R AML	AML	AML	R/R AML	MDS/MPN	MDS/MPN	sAML	R/R AML
ECOG	1	1	1	1	2	2	1	0	1
No. previous therapies	0	3	0	0	1	0	2	1	5
Previous IDH1^{MUT} inhibitor	No	No	No	No	No	No	Yes	No	Yes
WBC at baseline ($10^3/\mu l$)	11	0.4	1.2	1.5	0.6	3.7	4.6	1.8	1.9
Hb at baseline (g/dl)	10.1	7	10.1	9.4	8.9	8	11.2	8.7	7.7
Platelets at baseline ($10^3/\mu l$)	53	12	6	185	5	57	160	74	12
Blast Count at baseline (%)	28	35	79	47	61	16	12	10	23
IDH1 mutation at baseline (VAF [%])	R132H (29)	R132C (23)	R132C (42)	R132H (22)	R132C (31)	R132H (32)	R132H (38)	R132H (17)	R132S (18)
Treatment	IVO 500mg + VEN 800mg	IVO 500mg + VEN 400mg + AZA 75mg/m^2	IVO 500mg + VEN 400mg + AZA 75mg/m^2	IVO 500mg + VEN 400mg	IVO 500mg + VEN 800mg	IVO 500mg + VEN 800mg	IVO 500mg + VEN 400mg	IVO 500mg + VEN 400mg + AZA 75mg/m^2	IVO 500mg + VEN 800mg
Time on treatment (months)	3.9	6.4	26.8	47.7	5.5	9.5	13.7	49.0	58.2
Treatment failure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Duration of response (months)	3.4	5.4	24.3	47.8	3.9	6.7	12.6	48.1	57.0
Overall survival (months)	4.2	7.4	31.3	62.0	8.8	49.7	38.1	49.0	58.2
Survival status	Deceased	Deceased	Deceased	Alive	Deceased	Alive	Deceased	Alive	Alive

Table 3.1: Patient characteristics. Clinical characteristics for the patient cohort. ECOG: Eastern Cooperative Oncology Group; Hb: haemoglobin; WBC: white blood cell count; IVO: ivosidenib; VEN: venetoclax; AZA: azacitidine. Clinical data was collected by colleagues at the MD Anderson Cancer Center.

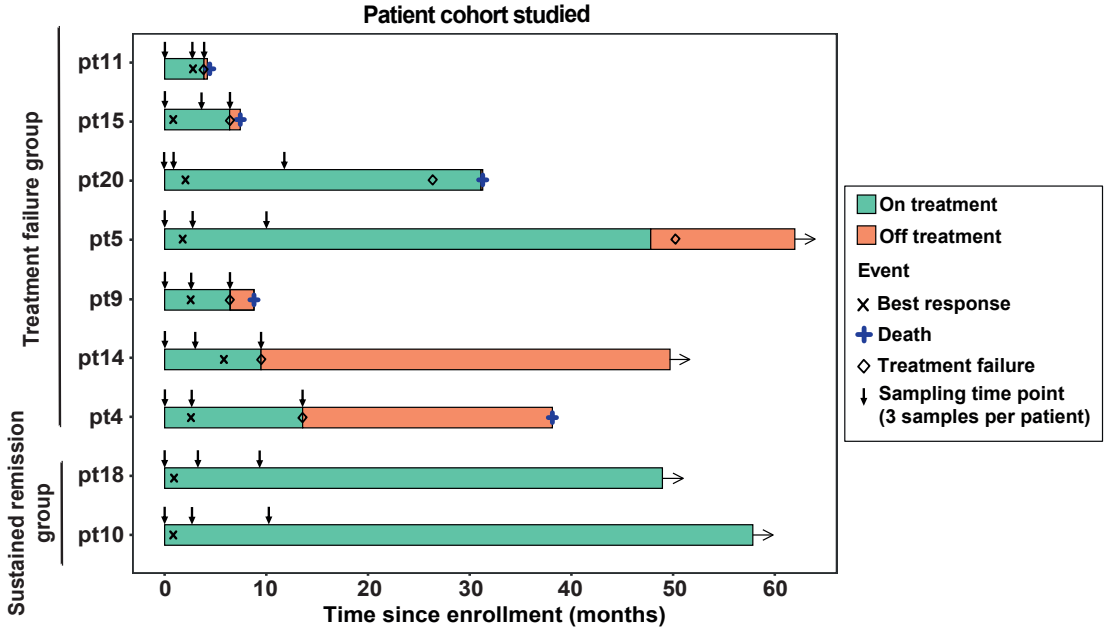


Figure 3.1: Patient cohort studied and treatment responses. Swimmer plot showing the response to treatment for the 9 studied patients. X indicates the time point when the best response was first achieved. Treatment failure (relapse, progression, or necessity to change treatment) is indicated by the black rhombus. Vertical arrows indicate the time points of sampling.

For each patient, 3 sequential samples were analysed: the first sample was collected at baseline; the second was collected either when the best response was first documented, or shortly before/after best response; the third was collected either from relapse (for pts 4/9/11/14/15) or from longer-term remission (for pts 5/10/18/20). The median duration of response (DOR) was 12.7 months; the median time on treatment was 13.7 months; the median overall survival (OS) was 38.5 months. At the time of data cut-off (12/14/2023), 2 patients (pts 10/18) remained on treatment.

3.2.2 Longitudinal immunophenotypic analysis of the patient cohort

I first characterized the immunophenotypic profiles for each sequential patient sample, to answer three questions: (a) From which BM compartments did the disease relapse? (b) How did treatment affect the BM cellular milieu? (c) Did abnormal sizes of immunophenotypic LSCs persist at early remission in patients experiencing rapid relapse?

To this end, multi-parameter flow cytometry analysis was performed (Figure 3.2). The FACS panel first separates live, single, Lineage$^-$ (Lin$^-$) MNCs. Mature erythroid and lymphoid cells are excluded by the Lin cocktail (detailed in section 2.5). CD34 and CD117 markers then distinguish immature CD34$^+$ HSPCs, Lin$^-$CD34$^-$CD117$^+$ GM precursors, and the more mature myeloid Lin$^-$CD34$^-$CD117$^-$ cells (hereafter referred to as mature myeloid cells). For granular inspection of immature populations, HSPCs are further segregated into HSC (Lin$^-$CD34$^+$CD38$^-$CD10$^-$ CD90$^+$CD45RA$^-$), MPP (Lin$^-$CD34$^+$CD38$^-$CD10$^-$ CD90$^-$CD45RA$^-$), LMPP (Lin$^-$CD34$^+$CD38$^-$ CD10$^-$CD45RA$^+$), MLP (Lin$^-$CD34$^+$CD38$^-$CD10$^+$), CMP (Lin$^-$CD34$^+$CD38$^+$CD10$^-$ CD123^{mid}CD45RA$^-$), GMP (Lin$^-$CD34$^+$CD38$^+$ CD10$^-$ CD123$^{mid/hi}$CD45RA$^+$), MEP (Lin$^-$CD34$^+$CD38$^+$ CD10$^-$CD123$^-$CD45RA$^-$), and B-NKp (Lin$^-$CD34$^+$CD38$^+$CD10$^+$) sub-compartments. Given that BM population sizes are abnormal in myeloid malignancy, FACS gates were first set using a normal healthy control BM sample (Figure 3.2a). The same gates were then used to analyse patient samples (Figure 3.2b). To establish the sizes of normal BM cell



Figure 3.2: Representative FACS panels. **(a)** Representative FACS panel from normal human BM. Sizes of each immunophenotypic population as percentage of all live, Lin$^-$ cells, are shown. FACS definitions of each population are listed in the main text. **(b)** As in (a) but showing a representative patient sample from trial entry (pt18), with expanded LMPP and GMP.

compartments, I first analysed 10 healthy age- and sex-matched BM donors, followed by analysis of patient samples (Table 3.2 and Figures 3.3 - 3.4).

Whilst HSPC and/or GM precursor compartments were expanded at diagnosis in most patients, treatment rapidly re-established myeloid maturation in all patients

			% of Lin-			% of HSPC							
Sample	Gender	Age	CD34+	Precursors	Mature	HSC	MPP	LMPP	CMP	GMP	MEP	MLP	B-NK
NOC025	Male	70.3	12	4	84	2.3506851	3.61881785	0.11341025	41.2400895	30.2083656	19.0735414	0.05464312	3.34044725
NOC027	Female	65.7	15	2	83	7.99540134	4.60911371	0.30309365	38.8795987	23.6204013	22.5752508	0.14632107	1.8708194
NOC045	Female	78.5	12	4	84	3.60041623	1.7169615	0.12486993	38.1893861	35.3798127	19.7710718	0	1.21748179
NOC070	Male	65.9	16	3	81	9.8162029	5.40096519	0.1950919	45.8979361	12.8349933	25.5673067	0	0.28750385
NOC076	Male	68.4	9	2	89	6.25196356	2.99507802	0.31416902	34.1397005	26.5996439	25.3429679	0.0628338	4.29364331
NOC088	Female	70.1	6	3	91	10.1380637	12.3686492	1.36372286	33.8287841	20.8258452	20.7201302	0.03594308	0.71886166
NOC094	Female	75.5	9.5	2.5	88	12.0963117	5.69360877	0.59438773	31.2835647	23.8797877	18.2487461	0.03858306	8.16501038
NOC095	Male	75.3	5	1	94	9.29743285	7.80984359	0.89255355	20.4012241	24.1201972	32.0894254	0.04462768	5.34469568
NOC156	Female	63.1	18	3	79	5.89491109	4.21973477	0.34816982	40.9360139	24.622069	12.4361254	0.17220854	11.3707675
NOC186	Female	80.9	6	2	92	10.1186103	8.95350058	0.79773276	32.5390994	20.8880025	25.4014905	0.04198593	1.25957804
		Average	**10.85**	**2.65**	**86.5**	**7.75599987**	**5.73862731**	**0.50472015**	**35.7335397**	**24.2979118**	**22.1226056**	**0.05971463**	**3.78688089**

Table 3.2: FACS analysis of normal BM. Samples are derived from the MARCH cohort [80]. The values from the NOC156 sample are averaged across 10 technical replicates.

(Figure 3.3a), consistent with all patients initially achieving a clinical response. Rapid induction of myeloid maturation occurred regardless of DOR, although there was a trend towards more immature cells early during treatment in patients with shorter DOR (Figure 3.3b). In 4/5 patients from which a relapse sample was available, differentiation arrest of immature cells was re-established at relapse (Figure 3.3c). This excludes that monocytic differentiation, associated with venetoclax resistance [234, 368], was a predominant resistance mechanism in this cohort.

Next, HSPC sub-compartments were analysed. Expectedly, HSPC ratios were perturbed both at diagnosis and relapse (Figure 3.4a). Across patients, MPP/LMPP/GMP/CMP compartments were expanded at baseline and/or relapse, consistent with prior findings that these sub-compartments are frequently expanded in AML [100, 240]. Patterns of expansion differed between patients, but sometimes even between baseline and relapse within the same patient (specifically pronounced in pts 15/14). At an early treatment time point, HSPC sub-compartment ratios were near normalized in some patients (pts 9/14/20/5/18/10), but not in others (pts 11/15/4). Interestingly, the contribution of MEPs early during treatment was significantly lower in patients with shorter DOR (Figure 3.4b), indicating a less efficient re-establishment of multi-lineage haematopoiesis in patients experiencing more rapid relapse. In patients for which a relapse sample was available, I analysed the longitudinal variation in those populations that showed aberrant contribution to HSPCs at relapse (Figure 3.4c). Interestingly, in pts 11/15/4, the contribution of these populations to HSPCs early during treatment was higher compared to

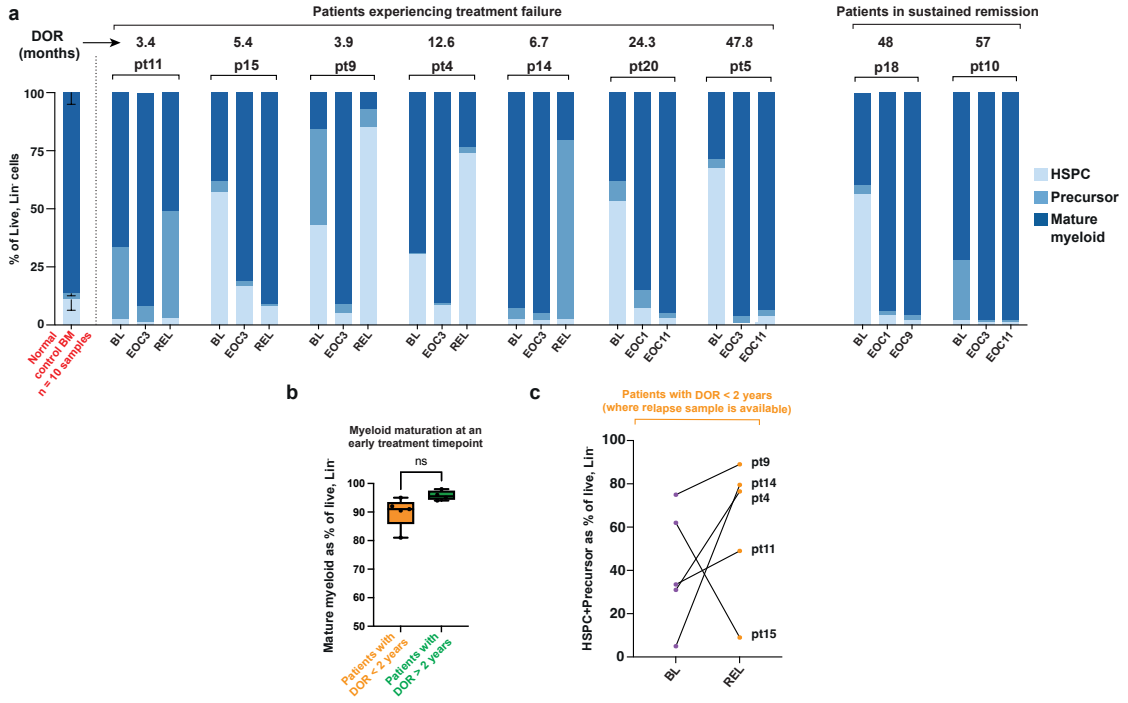


Figure 3.3: FACS analysis of HSPC, GM precursor, and mature myeloid compartments. **(a)** Immunophenotyping of HSPC, GM precursor, and mature myeloid populations in patient samples (BL, baseline; EOC, end of cycle; REL, relapse). Populations are quantified as percentage of all live, Lineage$^-$ (Lin$^-$) cells. As control, immunophenotyping was performed on healthy BM (n = 10 biologically independent samples; leftmost bar). Error bars indicate the standard deviation (sd) of the mean. DOR, duration of response. **(b)** Contribution of mature myeloid cells to all live, Lin$^-$ cells at an early treatment time point, comparing patients who achieved DOR $<$ or $>$ 2 years. Each dot represents a patient. Median values are shown. Whiskers extend from minimum to maximum value. Student's t-test was used. **(c)** Contribution of immature cells (HSPCs and GM precursors) to all live, Lin$^-$ cells at baseline versus relapse, for patients where the relapse sample was available.

normal BM. Moreover, in pts 11/15/14, the contribution of these populations to HSPCs increased from pre-treatment to early treatment response (Figure 3.4c), despite HSPCs not being aberrantly expanded at that point in time (Figure 3.3a). This suggested that, in some patients, cellular foundations of relapse were being established early during treatment. However, these patterns were not observed for pt9.

Collectively, cellular bases of the disease at diagnosis and relapse were heterogeneous. Myeloid maturation was induced early in all patients, consistent with a clinical response, regardless of patient outcome. Whilst ratios of HSPC

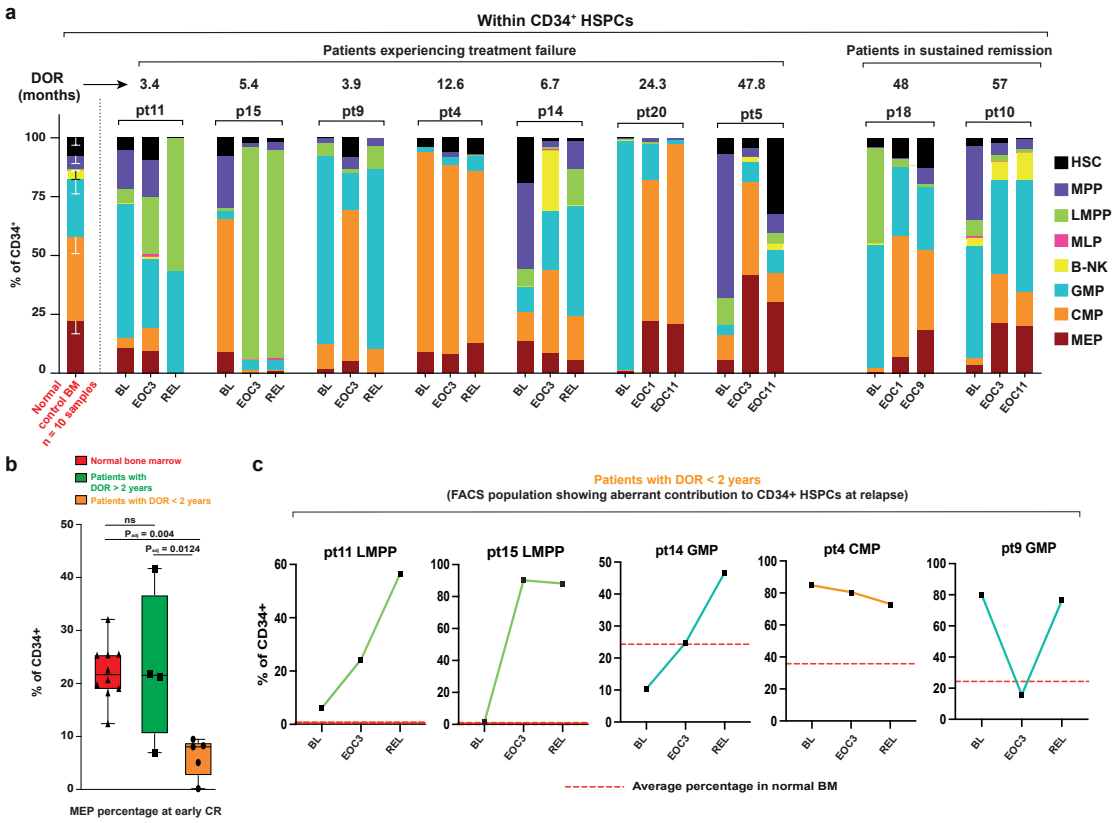


Figure 3.4: FACS analysis of HSPC sub-compartments. **(a)** HSPC sub-compartment composition in all patient samples and healthy BM controls. HSPC populations are quantified as percentage of all CD34$^+$ cells. The rest as in Figure 3.3a. **(b)** MEP contribution to all HSPCs at an early treatment time point, comparing normal BM and patients who achieved DOR $<$ or $>$ 2 years. Ordinary one-way ANOVA with Tukey's multiple comparisons test was used. **(c)** Contribution of a selected HSPC sub-compartment to all HSPCs in longitudinal patient samples for patients where the relapse sample was available. The red dotted line indicates the average size in normal BM.

sub-populations early during treatment were abnormal in some patients, no clear immunophenotypic correlates of treatment outcome were observed. However, robust conclusions in this regard require a larger patient cohort.

3.2.3 Application of TARGET-seq+ to sequential patient samples

To study the longitudinal variation in clonal structure underlying treatment failure or sustained response, I applied TARGET-seq+ to all sequential patient samples, to obtain clonal, immunophenotypic, and RNA-seq data at single-cell resolution (Figure 3.5).

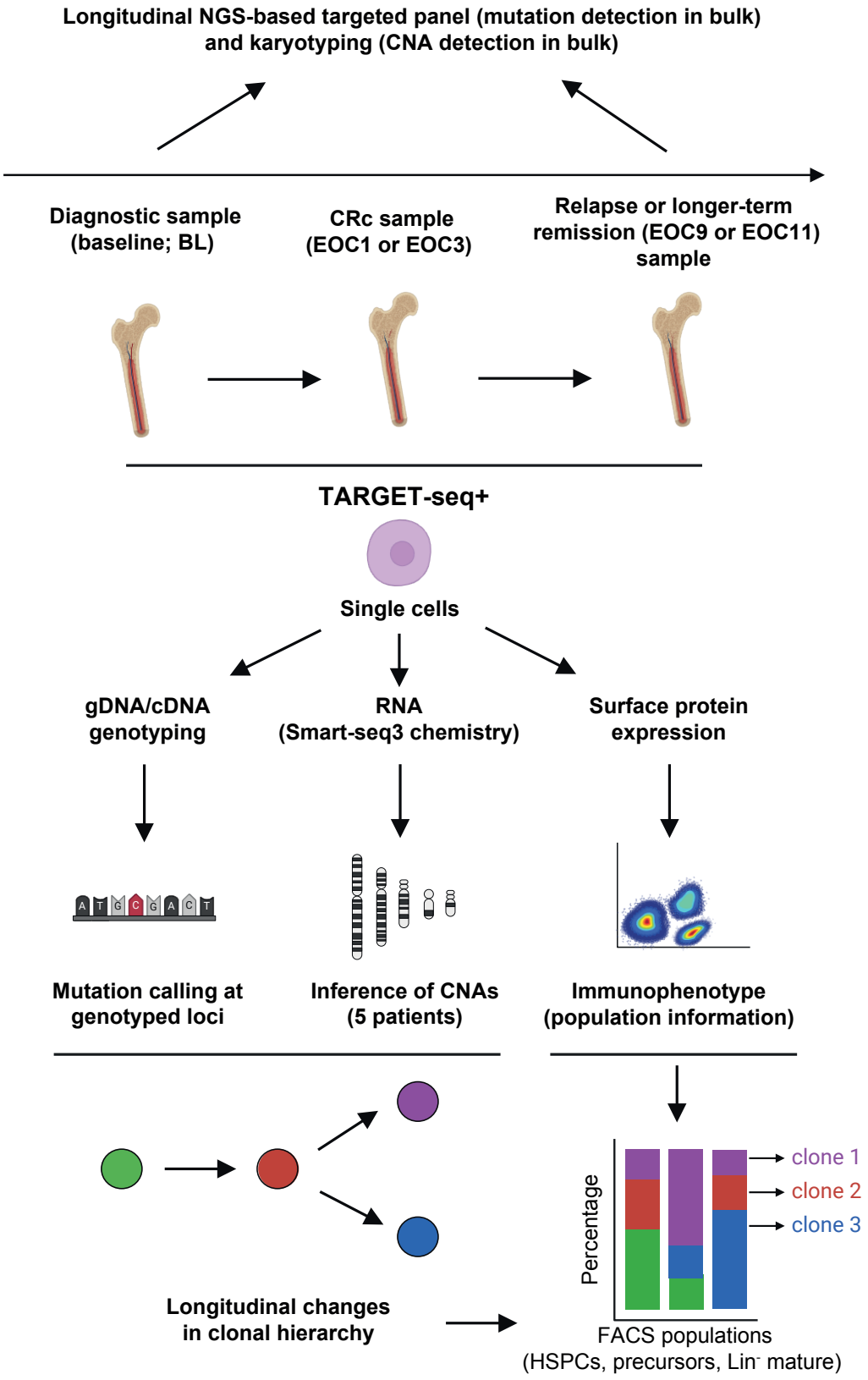


Figure 3.5: TARGET-seq+ framework for clonal, transcriptomic, and immunophenotypic single-cell analysis of sequential patient samples. For technical details, please refer to section 2.7 and Appendix A.2.

Patient	Karyotype at baseline	Karyotype at treatment failure
4	46,XY,t(4;15)(q12;q11.2)[1]/46,XY,del(20)(q11.2q13.3)[1]/46, XY[18]	46,XY,del(20)(q11.2q13.3)[5]/46,idem,+1,der(1;13)(q10;q10)[9]/46, idem,+1,der(1;15)(q10;q10)[3]/46,XY[3]
5	46,XX,del(5)(q22q35)[6]/46,idem,idic(22)(p11.2)[1]/46,XX[8]	46,XX,del(5)(q22q35)[4]/46,XX[7]
9	47,XY,del(5)(q22q33),+11,del(12)(p11.2p13)[9]/46,XY[11]	46,XY,del(12)(p11.2p13)[2]/46,idem,del(5)(q22q33)[12]/46,XY[6]
10	46,XX[5]//46,XY[7]	NA
11	46,XX[20]	46,XX[20]
14	47,XX,+4[1]/46,XX[19]	48,XX,+4,+21[14]/49,XX,+4,+20,+21[4]/46,XX[2]
15	46,X,t(X;9)(p11.2;q34),del(5)(q11.2q33),add(7)(p11.2),dup(21) (q21q22)[19]/46,XX,del(1)(q12),del(5)(q11.2q33),add(7) (p11.2),add(9)(q34),dup(21)(q21q22)[1]	46,XX,t(3;6)(q21;p21.3),del(5)(q11.2q33),add(7)(p11.2),dup(21) (q21q22)[7]/41~46,idem,add(3)(q29),+mar [cp13]
18	46,XY,del(7)(p13)[2]/46,XY[18]	NA
20	46,XY[20]	46,XY[20]

Table 3.3: Longitudinal karyotyping. Conventional karyotyping was performed for all patients at trial entry, and additionally at the time point of treatment failure, by colleagues at the MD Anderson Cancer Center. NA, not applicable.

Prior to TARGET-seq+, all patients were screened with NGS targeted panels interrogating recurrent myeloid somatic mutations (section 2.4) [80, 199], by colleagues at the MD Anderson Cancer Center and by Ms. Marlen Metzner at the University of Oxford. Patients were screened at baseline and, where applicable, at relapse. Moreover, a subset of patients was screened by WES, performed by Dr. Bilyana Stoilova at the University of Oxford (section 2.4). Collectively, these analyses identified recurrent somatic mutations within each patient, indicating the genomic loci to be genotyped by TARGET-seq+ (Tables A.1A.2). Moreover, all patients were karyotyped at baseline and, where applicable, at relapse, by colleagues at the MD Anderson Cancer Center (Table 3.3)[199].

Prior to TARGET-seq+, patient-specific genotyping primers (Tables A.3A.4) were tested first in a simulation of the TARGET-seq+ protocol on $CD34^+$ BM cells (exemplified in Figure 3.6), to identify genotyping primers which generate efficient genotyping amplicons (Figure 3.6a) whilst preserving cDNA libraries (Figure 3.6b).

By using FACS, I index-sorted single cells from the HSPC sub-compartments, GM precursor, and mature myeloid cell compartments from: (a) baseline; (b) an early therapy response; and (c) either treatment failure (for pts 4/9/11/14/15) or long-term remission (for pts 5/10/18/20) time points. Single-cell genotyping and RNA-seq libraries were generated using TARGET-seq+ [80]. For each cell, recurrent somatic mutations and copy number alterations (CNA; the latter identified by karyotyping and inferred in single cells from RNA expression) were called.

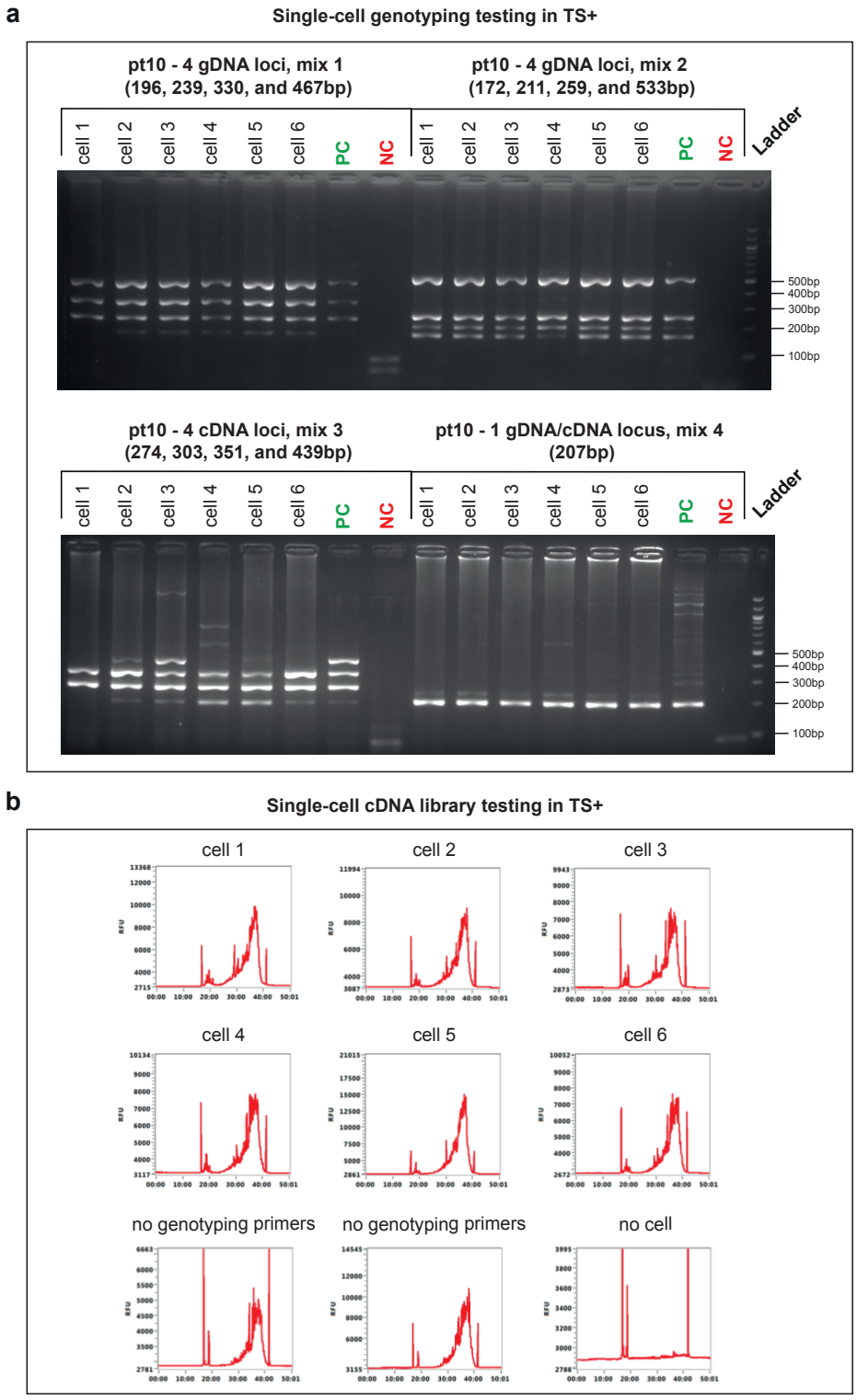


Figure 3.6: *(Legend on next page)*

Figure 3.6 ***(previous page)*****: Single-cell genotyping primer testing for TARGET-seq+ in human HSPCs. (a)** 2% agarose gel electrophoresis showing representative genotyping amplicons in single cells, when 4 PCR1 genotyping mixes were used. Expected amplicon sizes shown above. PC, positive control (K562 gDNA for genomic genotyping primers or K562 cDNA for cDNA genotyping primers); NC, negative control (empty well). **(b)** Representative single-cell cDNA libraries generated via TARGET-seq+. Cell 1 – 6 show libraries generated in presence of genotyping primers from (a). Two single-cell libraries were generated in absence of genotyping primers. An empty well was used as negative control.

Of note, I did not interrogate specific *CEBPA* (pts 11/9) and *RUNX1* (pt9) mutations, as I could not design suitable primers for single-cell genotyping. For pts 4/5/9/11/14/18, I did not genotype mutations with low VAFs ($\leq$ 5%) at relapse, hypothesizing they are likely to make a small contribution to changes in clonal dynamics with treatment. Mutational and CNA readouts were used to assign single cells to genetic clones (metrics shown in Table A.5) and patient-specific clonal hierarchies were reconstructed. Overall, 15,398 cells were assigned to a clone.

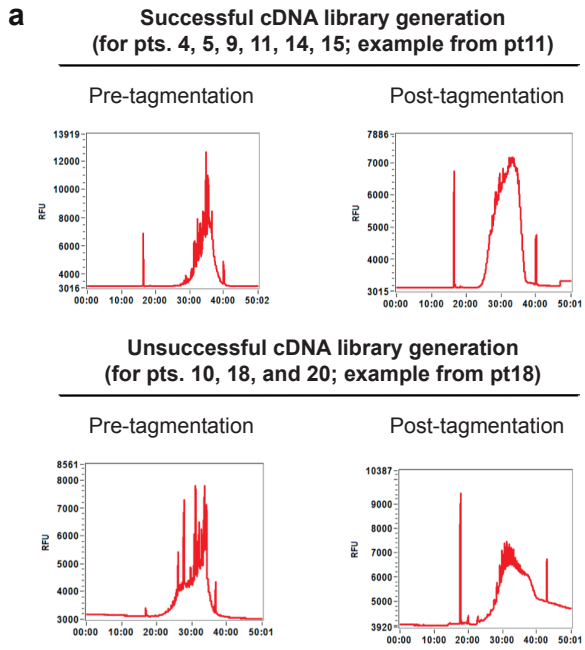


Figure 3.7: TARGET-seq+ libraries generated from a patient sample. Successful (top) and failed (bottom) cDNA (left) and tagmented (right) libraries from single cell pools.

Finally, I integrated the clonal status and immunophenotypic profile for each cell,

within each sample (Figure 3.5). Of note, cDNA libraries were successfully generated for pts 4/5/9/11/14/15, but failed for pts 10/18/20, due to TSO concatemers (Figure 3.7). However, longitudinal karyotyping in pts 10/18/20 did not reveal cytogenetic evolution throughout treatment; hence, CNA inference from RNA-seq was not critical in these patients.

3.3 Newly detected genetic clones are selected within LSCs months or years prior to treatment failure

In 7 patients who experienced treatment failure (pts 4/5/9/11/14/15/20), there were 3 different patterns of disease progression. 4 patients (pts 11/15/5/20) relapsed with clones harboring recurrent genetic driver mutations undetected prior to treatment. Within this group, 2 patients (pts 11/15) relapsed rapidly after achieving CR and 2 patients (pts 5/20) relapsed after years of response. In the remaining 3 relapsing patients (pts 14/9/4), relapse-associated clones were detected already at trial entry but were not the dominant clones. Upon therapy exposure, these clones rapidly lost the capacity to differentiate into mature myeloid cells and ultimately became dominant at relapse. Intriguingly, rapid selection of resistant clones was universally observed within 3 cycles of therapy, regardless of DOR. I first analyzed those patients who relapsed with newly detected genetic clones.

3.3.1 *FLT3*-ITD or *NF1*-mutant clones emerge within LSCs 3 cycles into treatment and associate with rapid relapse

2 patients (pts 11/15) had short DORs (3.4 and 5.4 months, respectively), with therapy-induced selection of genetically evolved clones, not detected at trial entry, but which were present by end of cycle 3 (EOC3).

In pt11, somatic *DNMT3A* (D), *NPM1* (N), *IDH1* (I), and *NRAS* (N) mutations were detected pre-treatment (genotyping thresholds in Figures A.6a-b). The two major clones were DNI and DNIN (Figures 3.8a-b and Table A.6). Both clones dominated the GM precursor-like compartment, which was the most expanded BM population, but were also detected within LMPP-like and GMP-like populations (Figure 3.8c). All three compartments harbor LSC potential in AML [100, 110].

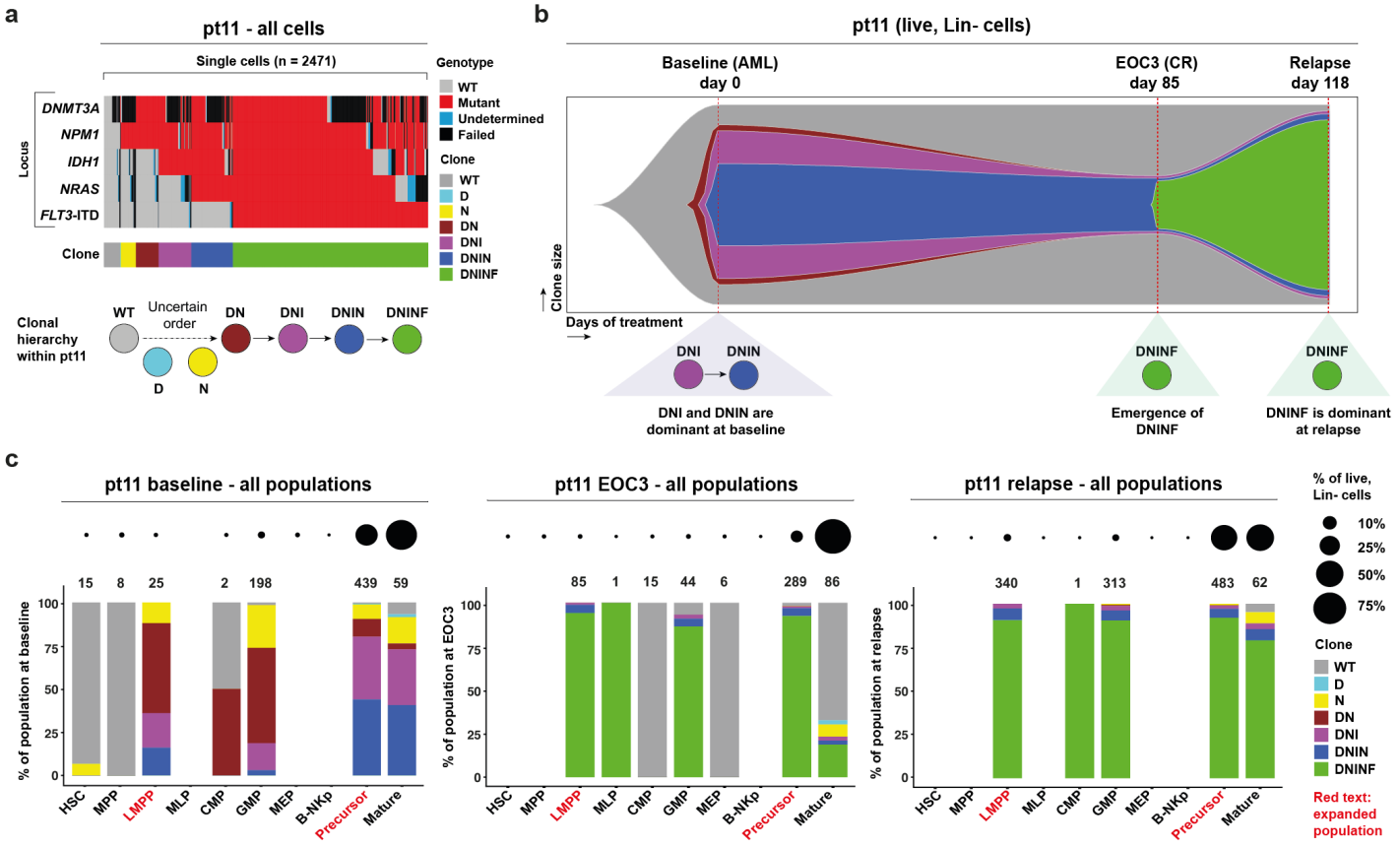


Figure 3.8: Longitudinal clonal analysis for pt11. (a) Top, single-cell genotyping for pt11, within cells assigned to clones (n = 2,471). Columns represent single cells. For each locus (mutations as in Figure A.6), cells are either wild type (WT, gray), mutant (red), undetermined (blue), or failed (insufficient coverage, black). Middle, clonal assignment for each cell (legend on the right). Due to suboptimal *DNMT3A* genotyping, the order of acquisition of the *DNMT3A* and *NPM1* mutations was uncertain. Clone abbreviations: WT: wild-type (gray). D: *DNMT3A*-mutant (cyan). N: *NPM1*-mutant (yellow). DN: *DNMT3A*/*NPM1*-mutant (brown). DNI: *DNMT3A*/*NPM1*/*IDH1*-mutant (purple). DNIN: *DNMT3A*/*NPM1*/*IDH1*/*NRAS*-mutant (blue). DNINF: *DNMT3A*/*NPM1*/*IDH1*/*NRAS*/*FLT3*-mutant (green). Bottom, most likely clonal structure for pt11. **(b)** Longitudinal changes in clone size within all live, Lin$^-$ cells (HSPC, GM precursors, and mature myeloid cells) in pt11. Clone colour-coding as in (a). The x-axis is proportional to the time relative to baseline. Clone size inference detailed in section 2.8.7. **(c)** Clonal composition of each sorted immunophenotypic population at each time point in pt11. Clone colour-coding as in (a). Numbers of single cells analysed per population are indicated above. The size of the black dot above each bar (legend on the right) is proportional to the percentage of that population within all live, Lin$^-$ cells (defined by FACS). The red text marks populations exceeding the mean size of the respective healthy control BM populations by more than 4 × SD.

By EOC3, though pt11 had achieved CR and most mature myeloid cells were WT, the DNIN clone evolved by acquiring an internal tandem duplication within the *FLT3* gene (*FLT3*-ITD; F), resulting in the DNINF clone (Figures 3.8a-b). At EOC3, the DNINF clone comprised 24% of all live, Lin^- cells, but ≈90% of the LMPP-like, GMP-like, and GM precursor-like populations, none of which substantially expanded at that point (Figure 3.8c). Concordant with the highly proliferative phenotype associated with *FLT3*-ITD, relapse ensued just 33 days following EOC3. At relapse, the DNINF clone was dominant throughout the whole spectrum of myeloid differentiation, from LMPPs down to more mature myeloid cells (Figure 3.8c). These observations are consistent with early therapy-induced selection of the relapse-causing *FLT3*-ITD clone within small LSC compartments, while the patient was in clinical remission. Of note, given the size of the DNINF clone at EOC3, it is likely that selection of DNINF occurred even earlier during treatment.

In pt15, genetically evolved clones were also selected within 3 cycles of treatment. Pre-treatment, mutations in *TP53* (T), *SF3B1* (S), and *IDH1* (I) were detected (genotyping thresholds in Figures A.7a-b). Additionally, del(5)(q11.2q33) was detected by cytogenetics (Table 3.3) and was inferred at single-cell level from gene expression at chromosome 5 (abbreviated (-5q); Figures A.7c-d and A.1). The T(-5q)SI clone was dominant pre-treatment within multiple expanded HSPC compartments and only partially completed myeloid differentiation (Figures 3.9a-c). T(-5q)SI was eradicated by EOC3, when pt15 attained CR with incomplete haematologic recovery (CRi; Figures 3.9a-c). Simultaneously, the *IDH1*-WT ancestral T(-5q)S clone had acquired 2 consecutive loss-of-function (LoF) mutations in the *NF1* (N) tumor suppressor [369–371] (T(-5q)SN and T(-5q)SNN clones). At EOC3, *NF1*-mutant clones contributed to 17/32 (53%) of sorted LMPPs (Figures 3.9a-c). Further expansion of *NF1*-mutant clones associated with clinically overt relapse, ≈3 months following EOC3. Of note, whilst multiple progenitor compartments were expanded at baseline, only the LMPP was expanded at relapse, suggesting distinct clones can lead to expansion of different BM compartments.

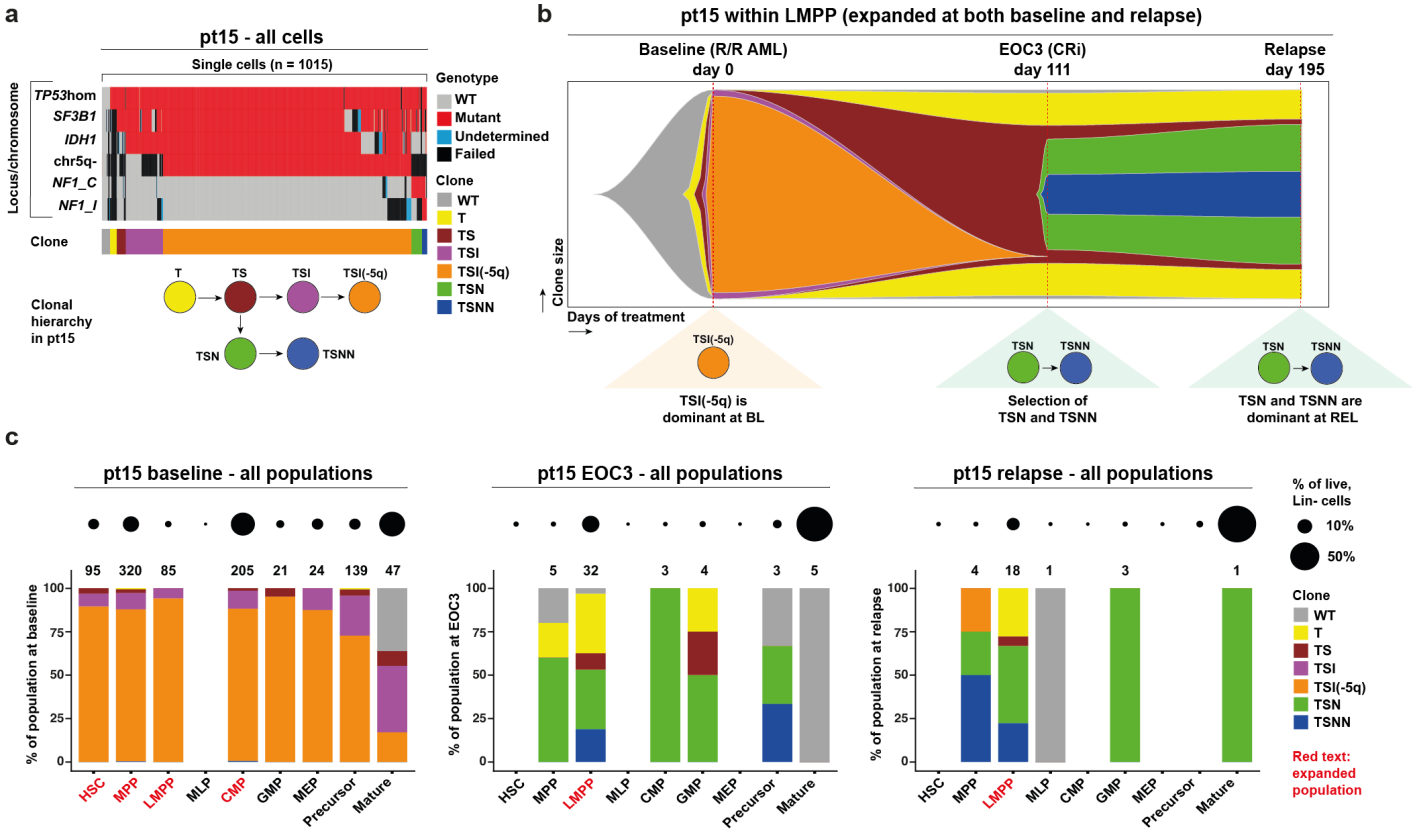


Figure 3.9: Longitudinal clonal analysis for pt15. (a) Same as in Figure 3.8a but for pt15. Somatic aberrations as in Figures A.7a-c. Clone abbreviations: WT: wild-type (gray). T: *TP53*-mutant (yellow). T(-5q): *TP53*-mutant with chr5q- (brown). T(-5q)S: *TP53/SF3B1*-mutant with chr5q- (purple). T(-5q)SI: *TP53/SF3B1/IDH1*-mutant with chr5q- (orange). T(-5q)SN: *TP53/SF3B1/NF1*-C2223*-mutant with chr5q- (blue). T(-5q)SNN: *TP53/SF3B1/NF1*-C2223*/*NF1*-I679fs-mutant with chr5q- (blue). **(b)** Same as in Figure 3.8b but for pt15, within the LMPP. Clone colour-coding as in (a). **(c)** Same as in Figure 3.8c but for pt15. Clone colour-coding as in (a).

Again, the size of *NF1*-mutant clones at EOC3 suggests resistant clones were selected even earlier during treatment.

Overall, in both patients, resistant clones, undetected at diagnosis, were selected rapidly within the first treatment cycles. Selection occurred at remission, within small immunophenotypic LSC populations.

3.3.2 Resistant clones emerge years prior to treatment failure and expand slowly in a competing clonal milieu

In 2 other patients (pts 5/20), newly detected clones were also selected early. However, these patients had longer DOR (Figure 3.1; Table 3.1), and frank treatment failure ensued with substantially slower kinetics.

Pt20 had a DOR of 24.3 months. Pre-treatment, two *TET2* (T), and *ASXL1* (A), *SRSF2* (S), *IDH1* (I), and *BRAF* (B) mutations (genotyping thresholds in Figure A.8) were detected within a complex clonal structure (Figures 3.10a-b). The TTASI clone was dominant within the expanded GMP-like compartment and differentiated poorly into mature myeloid cells (Figure 3.10c). By EOC1, when BM blasts were reduced from 79% to 6%, the TTASI clone was eradicated and replaced by multi-clonal hematopoiesis from clones ancestral to the TTASI clone (TTA, TTAB, and TTAS), in the progenitor, GM precursor, and mature compartments (Figures 3.10b-c). In parallel, a newly detected *RUNX1*-Y414fs*186 mutation was selected within a minor TTASR clone. At EOC1, TTASR competed with a larger *BRAF*-mutant TTAB clone. However, by EOC11, when pt20 was in morphological leukemia-free state (MLFS), the TTASR and TTAS clones had outcompeted TTAB in the immature HSPC compartments, whereas TTAB and the ancestral TTA and TT clones were more prominent in mature myeloid cells (Figure 3.10c). 15 months later, pt20 progressed to AML, with expansion of a primitive myeloid population and the *RUNX1*-Y414fs* mutation at a VAF of 69% in BM MNCs (Table A.2) [199], consistent with the hypothesis that defective differentiation of the TTASR clone ultimately led to treatment failure.

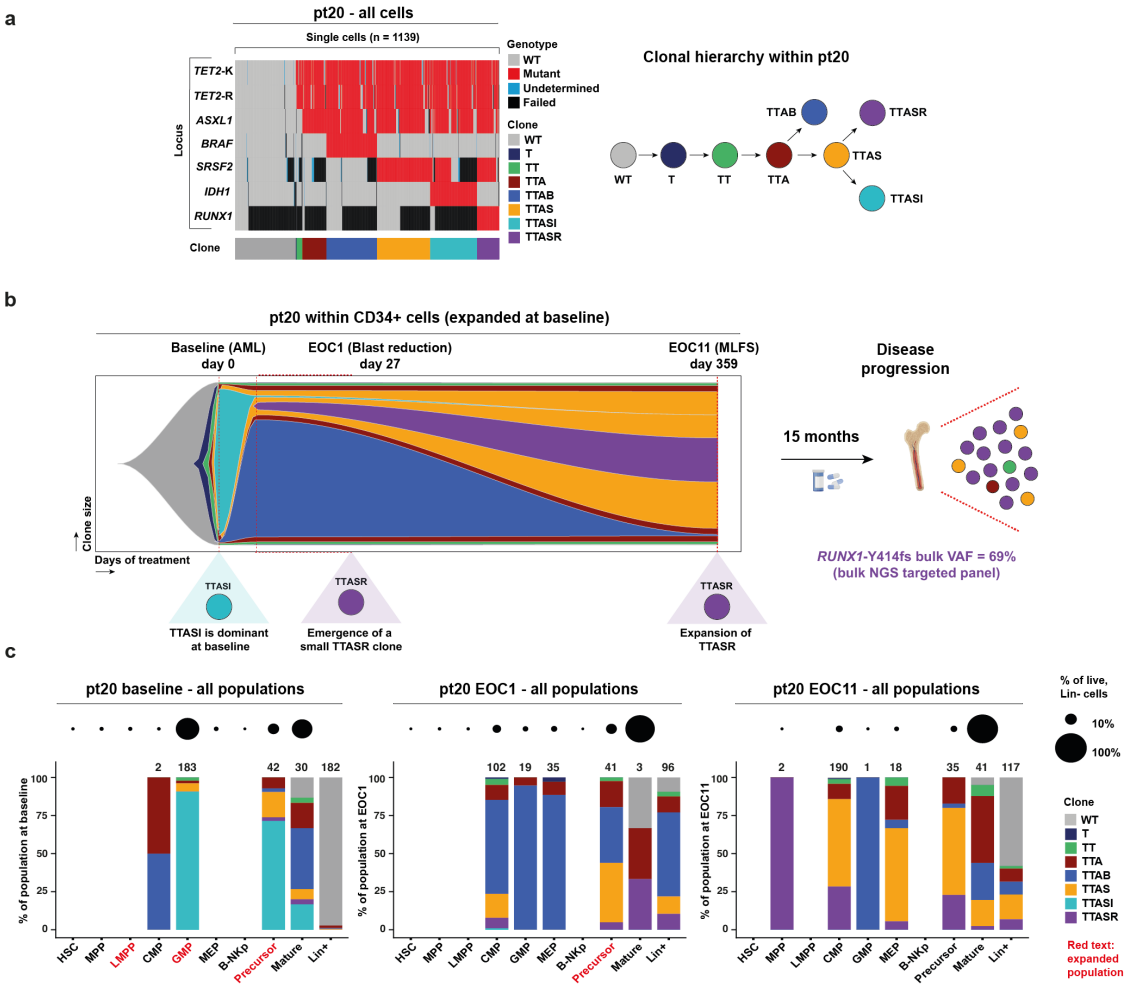


Figure 3.10: Longitudinal clonal analysis for pt20. **(a)** Same as in Figure 3.8a but for pt20. Mutations as in Figure A.8. Clone abbreviations: WT: wild-type (gray). T: *TET2*-K-mutant (dark blue). TT: *TET2*-K/*TET2*-R-mutant (green). TTA: *TET2*-K/*TET2*-R/*ASXL1*-mutant (brown). TTAS: *TET2*-K/*TET2*-R/*ASXL1*/*SRSF2*-mutant (orange). TTASI: *TET2*-K/*TET2*-R/*ASXL1*/*SRSF2*/*IDH1*-mutant (cyan). TTAB: *TET2*-K/*TET2*-R/*ASXL1*/*BRAF*-mutant (blue). TTASR: *TET2*-K/*TET2*-R/*ASXL1*/*SRSF2*/*RUNX1*-mutant (purple). **(b)** Same as in Figure 3.8b but for pt20, within HSPCs. Clone colour-coding as in (a). Right, 15 months following EOC11, pt20 progressed and the bulk VAF of *RUNX1*-Y414fs*186 reached 69%. **(c)** Same as in Figure 3.8c but for pt20. Clone colour-coding as in (a).

Of note, given that pt20 transitioned to ivosidenib monotherapy at EOC2, the outgrowth of the *IDH1*-WT TTASR clone was surprisingly slow and may be due to competition with TTAS and TTAB clones during treatment.

Pt5 responded to therapy for ≈4 years. Pre-treatment, del(5)(q22q35) ((-5q)), *RUNX1* (R), *ASXL1* (A), *IDH1* (I), and *TET2* (T) mutations were detected (genotyping thresholds and inferCNV analysis in Figures A.9a-d and A.2). The (-5q)RAI clone, dominant at baseline, was eradicated by EOC3 (Figures 3.11a-c and Table A.7), when the patient was in CR and most mature myeloid cells were WT (Figure 3.11c). Nevertheless, at EOC3, haematopoiesis was not fully devoid of somatic lesions. The ancestral (-5q) clone, which persisted in ≈1% of all BM Lin^- cells, acquired a previously undetected *SF3B1*-K700E mutation ((-5q)S clone) in < 0.1% of all BM Lin^- cells (Figures 3.11a-b). Interestingly, despite its small size, the (-5q)S clone was detected in 4/12 of sorted immunophenotypic HSCs (Figure 3.11c). In parallel, independent *TET2*-mutated CH (T clone) also persisted from baseline to EOC3. By EOC11, when pt5 was still in CR, all 3 clones (T, (-5q), and (-5q)S) had expanded (Figure 3.11b). However, whereas the T clone differentiated efficiently into mature cells, (-5q) and (-5q)S clones made a smaller contribution to mature cells, instead being clonally dominant in HSC (70% within HSCs). 40 months following EOC11, pt5 developed MDS with del(5)(q22q35) and *SF3B1*-K700E, and transitioned to lenalidomide monotherapy (Table 3.3 and Table A.2). Both del5(q) and *SF3B1*-K700E are recurrent in MDS [243, 372], where HSCs are frequently expanded and harbour LSC potential [242]. Hence, resistance-associated clones emerged ≈4 years prior to treatment failure but competed with *TET2*-mutated CH during years of remission.

Collectively, selection of newly detected clones can occur years prior to treatment failure, specifically within small immature BM compartments.

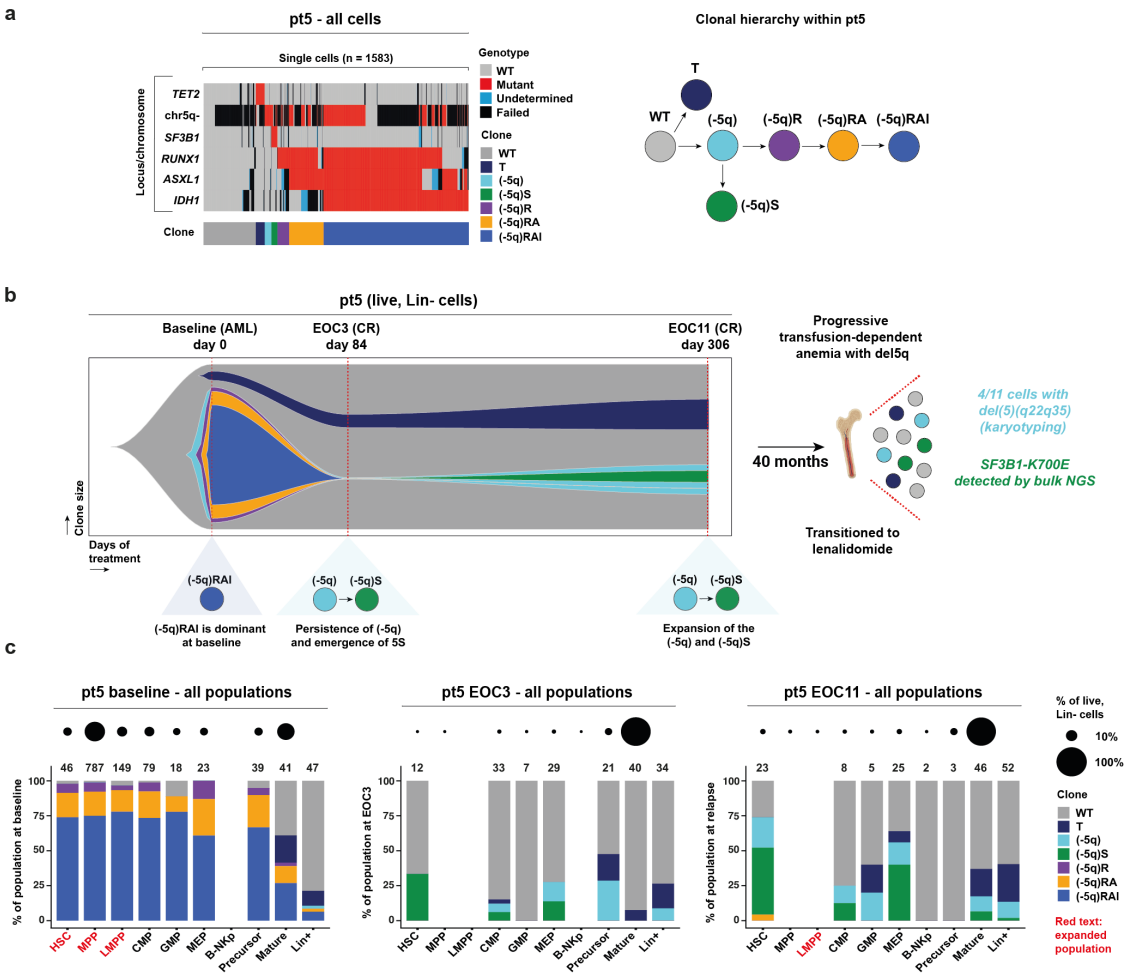


Figure 3.11: Longitudinal clonal analysis for pt5. (a) Same as in Figure 3.8a but for pt5. Somatic aberrations as in Figure A.9. Clone abbreviations: WT: wild-type (gray). T: *TET2*-mutant (dark blue). (-5q): chr5q- (cyan). (-5q)S: chr5q- and *SF3B1*-mutant (green). (-5q)R: chr5q- and *RUNX1*-mutant (purple). (-5q)RA: chr5q-, *RUNX1*/*ASXL1*-mutant (orange). (-5q)RAI: chr5q-, *RUNX1*/*ASXL1*/*IDH1*-mutant (blue). **(b)** Same as in Figure 3.8b but for pt5. Baseline and two sequential CR samples are plotted. Right, 40 months following EOC11, pt5 developed MDS with del(5)(q22q35) and *SF3B1*-K700E. **(c)** Same as in Figure 3.8c but for pt5. Clone colour-coding as in (a).

3.4 Genetic clones detected at baseline undergo rapid therapy adaptation

In all patients studied above, apparent genetic clonal evolution established the clonal foundations of resistant disease. In 3 other patients (pts 14/9/4), therapy resistance likely arose from complex clonal adaptation to treatment.

3.4.1 A pre-exiting clone harbouring trisomy 4 loses differentiation capacity rapidly upon treatment exposure

At trial entry, pt14 harbored *DNMT3A* (D), *NPM1* (N), and *IDH1* (I) mutations, and trisomy 4 (+4, genotyping thresholds and inferCNV analysis in Figures A.10a-d and A.3). The DNI clone was dominant at baseline (Figures 3.12a-b and Table A.8), whereas the much smaller DN(+4) clone was detected at similarly low levels at MPP, GM precursor, and mature myeloid differentiation stages (Figure 3.12c). By EOC3, when pt14 achieved CR, the DNI clone was almost undetectable and there was a partial reconstitution of WT hematopoiesis (Figures 3.12b-c). However, the DN(+4) clone was now detectable in the LMPP compartment, and expanded 25-fold within the GM precursor compartment, despite the small size of both populations (Figure 3.12c). Although the DN(+4) clone was present in these immature populations at high frequency, its contribution to mature myeloid cells was very limited (Figure 3.12c), suggesting it differentiated less efficiently compared to diagnosis. Hence, differentiation of DN(+4) was rapidly impaired under therapeutic pressure. ≈6 months later, relapse ensued, and the DN(+4) clone dominated the expanded GM precursor compartment, which became the major BM population with differentiation arrest (Figure 3.12c). The DN(+4) clone was able to differentiate to mature myeloid cells, but the higher frequency of the DN(+4) clone in HSPCs and GM precursors, compared to more mature myeloid cells, was consistent with impaired differentiation of this clone (Figure 3.12c). Though two additional smaller clones evolved from DN(+4) by acquiring either trisomy 20 (DN(+4+20) clone), or trisomy 21 (DN(+4+21) clone), the dominant clone at relapse was DN(+4) (Figure 3.12b). Of note, bulk NGS at relapse did not show evidence for novel driver somatic mutations at high VAF (Table A.2). The most parsimonious explanation for these findings is that DNI clone had a higher fitness advantage compared to DN(+4) prior to therapy, but inhibition of mutant IDH1, and loss of the DNI clone, enabled the previously less fit DN(+4) clone to dominate. Additionally, therapy could have induced epigenetic reprogramming, decreasing the differentiation capacity of DN(+4) clone.

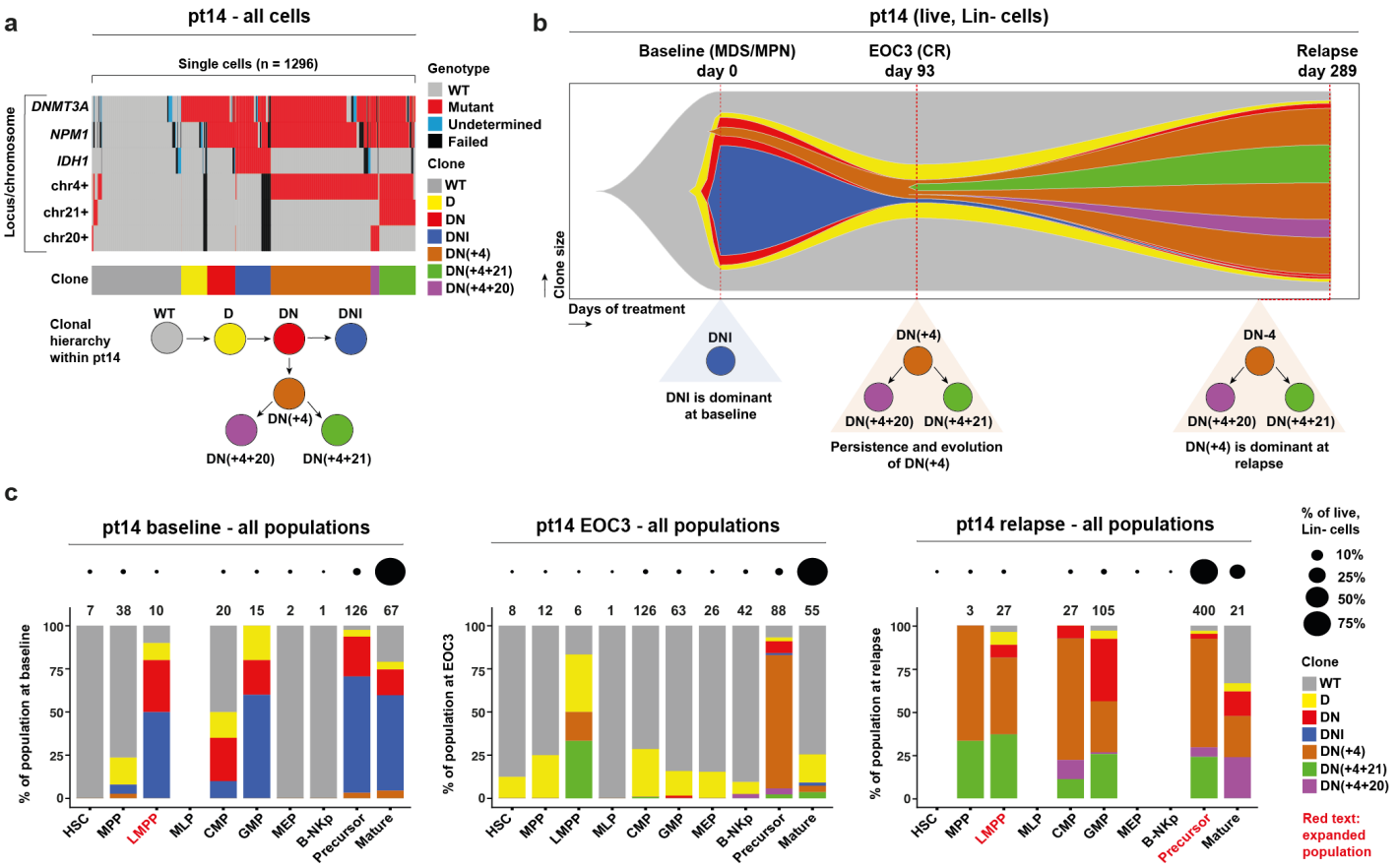


Figure 3.12: Longitudinal clonal analysis for pt14. (a) Same as in Figure 3.8a but for pt14. Somatic aberrations as in Figures A.10a-c. Clone abbreviations: WT: wild-type (gray). D: *DNMT3A*-mutant (yellow). DN: *DNMT3A*/*NPM1*-mutant (red). DNI: *DNMT3A*/*NPM1*/*IDH1*-mutant (blue). DN(+4): *DNMT3A*/*NPM1*-mutant with chr4+ (brown). DN(+4+21): *DNMT3A*/*NPM1*-mutant with chr4+ and chr21+ (green). DN(+4+20): *DNMT3A*/*NPM1*-mutant with chr4+ and chr20+ (purple). **(b)** Same as in Figure 3.8b but for pt14. **(c)** Same as in Figure 3.8c but for pt14. Clone colour-coding as in (a).

Of note, trisomy 4 is a rare event in myeloid malignancies, yet it has recently been associated with a hypermethylation phenotype, suggesting its possible role in ivosidenib resistance via chromatin mechanisms [373]. In this regard, a method allowing to study how somatic lesions influence chromatin regulation during treatment would be beneficial.

3.4.2 A *FLT3*-mutated clone rapidly loses differentiation capacity upon treatment exposure

In pt9, *ASXL1* (A), *IDH1* (I), *STAG2* (S), and *FLT3* tyrosine kinase domain (*FLT3*-TKD, F) mutations, del(5)(q22q33) (-5q), and trisomy 11 (11+), were detected at baseline (genotyping thresholds and inferCNV analysis in Figures A.11a-d and A.4). Pre-treatment, AIS and AI(-5q+11) clones were dominant (Figures 3.13a-b and Table A.9), but in different BM compartments (Figure 3.13c). A smaller *FLT3*-TKD-mutant clone (AISF) was present almost exclusively within GM precursors and was differentiating into more mature myeloid cells (Figures 3.13a-c). By EOC3, when the patient was in CR, most mature myeloid cells were either WT or harbored only an *ASXL1* mutation. However, AISF had gained clonal advantage within small immature HSPC compartments (MPP and LMPP and, to a lesser extent, HSC, CMP, and GMP) and differentiated inefficiently to mature myeloid cells (Figure 3.13c). In fact, whilst the AISF clone contributed to only 4% of all Lin^- BM cells at EOC3, it contributed to 64% of MPPs (54/84 cells), 60% of LMPPs (3/5 cells), and 26% of GMPs (8/31 cells). ≈3 months later, relapse ensued, with marked expansion of the GMP-like compartment, and, to a lesser extent, other progenitor populations and GM precursors. Progenitor- and GM precursor-like cells were now dominated by the AISF clone, which outcompeted *ASXL1*-single mutant cells, which were still dominant within HSCs. Compared to diagnosis, at relapse, the AISF clone differentiated much less efficiently into mature myeloid cells (Figure 3.13c). Of note, 2 *CEBPA* mutations and 1 *RUNX1* mutation were also detected by bulk NGS at relapse, at VAF > 10% (Table A.2), but I was unable to design efficient genotyping primers at these loci. This limits the dissection of clonal hierarchy at relapse.

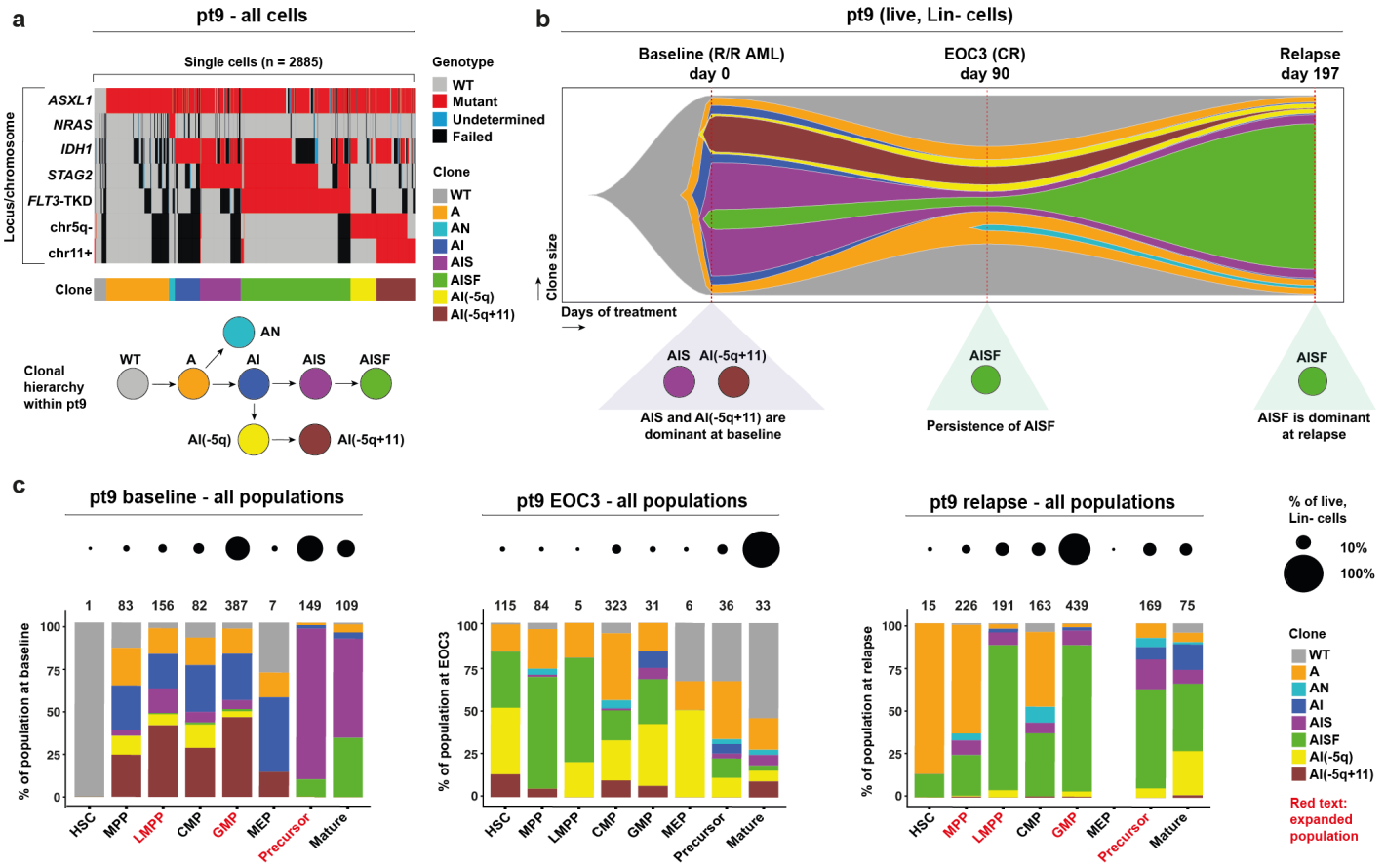


Figure 3.13: Longitudinal clonal analysis for pt9. (a) Same as in Figure 3.8a but for pt9. Somatic aberrations as in Figure A.11. Clone abbreviations: WT: wild-type (gray). A: *ASXL1*-mutant (orange). AN: *ASXL1*/*NRAS*-mutant (cyan). AI: *ASXL1*/*IDH1*-mutant (blue). AIS: *ASXL1*/*IDH1*/*STAG2*-mutant (purple). AISF: *ASXL1*/*IDH1*/*STAG2*/*FLT3*-mutant (green). AI(-5q): *ASXL1*/*IDH1*-mutant with chr5q- (yellow). AI(-5q+11): *ASXL1*/*IDH1*-mutant with chr5q- and chr11+ (brown). **(b)** Same as in Figure 3.8b but for pt9. **(c)** Same as in Figure 3.8c but for pt9. Clone colour-coding as in (a).

Nevertheless, these data are consistent with altered differentiation of the AISF clone after treatment exposure, which occurred either due to epigenetic mechanisms, genetic mechanisms, or a combination of both. Regardless of which mechanism was operative, rapid differentiation delay of AISF within LSCs occurred within 3 months of treatment, foreshadowing dismal outcome.

3.4.3 Pre-existing clones gradually lose differentiation capacity during treatment in a complex clonal milieu

Impaired differentiation of pre-existing clones also occurred in pt4 but was less marked. In this patient, who presented with MDS/MPN at baseline, there was a complex clonal architecture, and the BM was highly polyclonal throughout treatment. Somatic mutations in *SRSF2* (S), *IDH1* (I), *ASXL2* (A), *JAK2* (J), and *CBL* (C), gain of chromosome 1 (+1), loss of parts of chromosome 7 (-7), and del(20)(q11.2q13.3) (-20q), were detected pre-treatment (genotyping thresholds and inferCNV analysis in Figures A.12a-d and A.5). At baseline, CMPs, previously shown to harbor LSC activity [240], were expanded, but there was also a large population displaying mature myeloid immunophenotype. SIAJ and SIAC clones were dominant: SIAJ within HSPCs, and SIAC within more mature myeloid cells (Figures 3.14a-c and Table A.10). By EOC3, when the patient was in CR, the CMP compartment shrunk, likely contributing to the reduction in blast counts. The mature myeloid compartment comprised most cells, but still contained multiple abnormal clones (Figure 3.14c). At relapse, there was no single major dominant clone (Figure 3.14b). However, SIA(-20q+1) and SIA(-20q-7) clones, minor at baseline, expanded 2- and 69-fold, respectively, by EOC3, and showed the greatest increase in size at relapse (Figure 3.14b). Interestingly, both clones lost differentiation potential during treatment, and, by the time of relapse, accumulated within the re-expanded CMP-like compartment (Figure 3.14c). Of note, a mild increase of SIA(-20q+1) within CMPs was already detected at EOC3. del(20)(q11.2q13.3) is recurrent in myeloid malignancy [374] and trisomy 1 is observed in post-MPN AML [375].

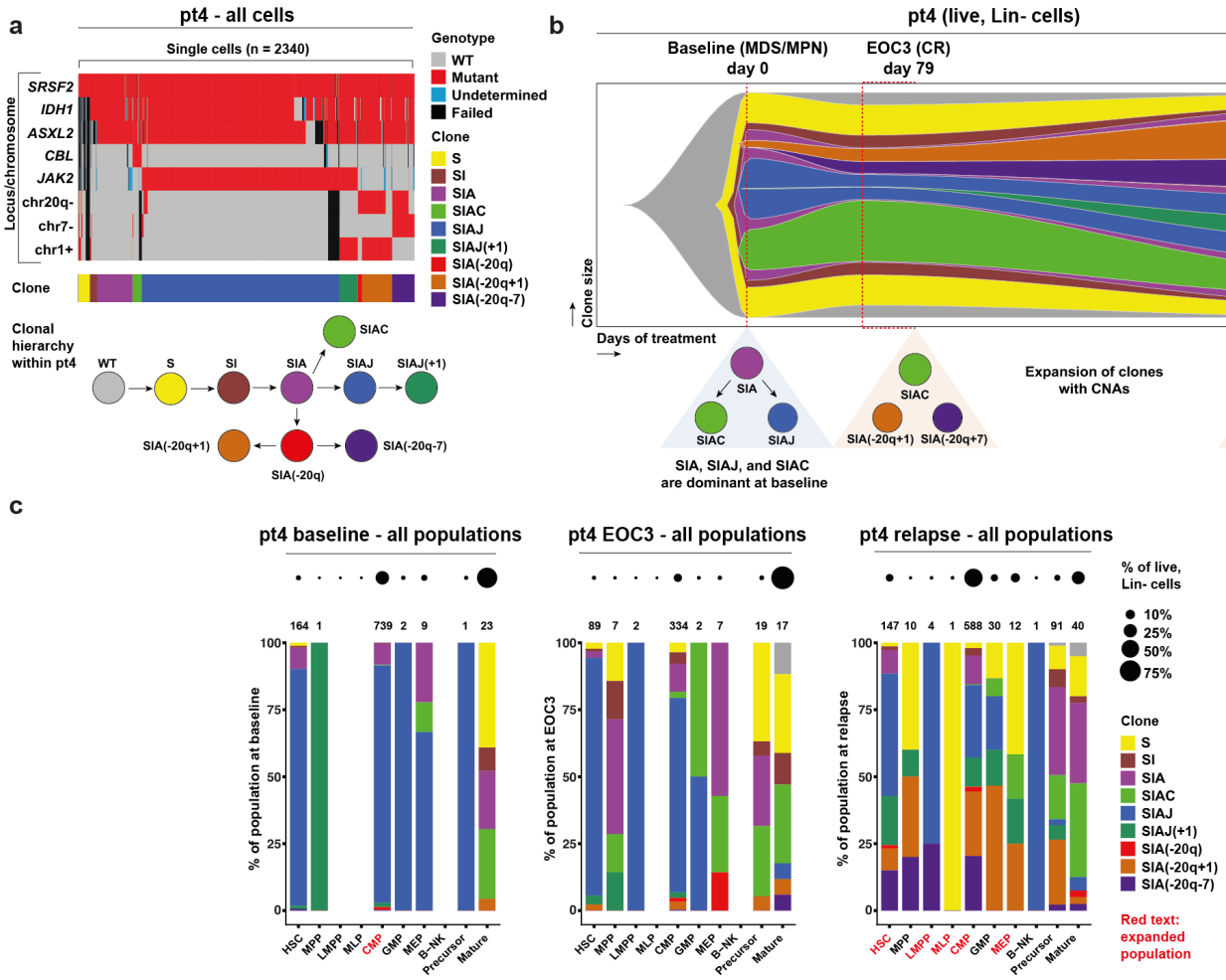


Figure 3.14: Longitudinal clonal analysis for pt4. **(a)** Same as in Figure 3.8a but for pt4. Somatic aberrations as in Figure A.12. Clone abbreviations: WT: wild-type (gray). S: *SRSF2*-mutant (yellow). SI: *SRSF2*/*IDH1*-mutant (brown). SIA: *SRSF2*/*IDH1*/*ASXL2*-mutant (light purple). SIAC: *SRSF2*/*IDH1*/*ASXL2*/*CBL*-mutant (light green). SIAJ: *SRSF2*/*IDH1*/*ASXL2*/*JAK2*-mutant (blue). SIAJ(+1): *SRSF2*/*IDH1*/*ASXL2*/*JAK2*-mutant with chr1+ (dark green). SIA(-20q): *SRSF2*/*IDH1*/*ASXL2*-mutant with chr20q- (red). SIA(-20q+1): *SRSF2*/*IDH1*/*ASXL2*-mutant with chr20q- and chr1+ (brown). SIA(-20q-7): *SRSF2*/*IDH1*/*ASXL2*-mutant with chr20q- and chr7- (dark purple). **(b)** Same as in Figure 3.8b but for pt4. **(c)** Same as in Figure 3.8c but for pt4. Clone colour-coding as in (a).

Moreover, monosomy 7 or partial chromosome 7 deletions are associated with adverse prognosis [172], poorer outcome following venetoclax treatment [376], and relapse following enasidenib (inhibitor of mutant IDH2) treatment [270].

In summary, impaired differentiation of pre-existing clones, either due to adaptive chromatin changes, or due to acquisition of unknown/undetected somatic mutations, was established within 3 cycles of treatment in pts 14/9/4. In fact, by EOC3, these clones accumulated within immunophenotypic LSC compartments.

3.5 Clonal foundations of sustained clinical response

Finally, 2 patients (pts 18/10) remained in remission for 4.1 and 4.8 years, respectively, at the point of data cut-off. Remission was marked by eradication of the leukaemic clones and expansion of both WT cells and CH. For both patients, I studied the pre-treatment sample and two sequential remission samples.

pt18 presented with sAML. *RUNX1* (R), *IDH1* (I), *SRSF2* (S), and *ASXL1* (A) somatic mutations (genotyping thresholds in Figures A.13a-b) were detected pre-treatment. Disease at baseline was driven by expanded LMPP- and GMP-like compartments, dominated by the RIS clone, and to a lesser extent RI clone (Figures 3.15a-c and Table A.11). The RIS clone did not differentiate efficiently into mature myeloid cells (Figure 3.15c). A minor independent *ASXL1*-mutated CH clone (A clone) was also detected (Figures 3.15a-b). By EOC1, when pt18 achieved CRi, the LMPP-like and GMP-like compartments shrank in size, whilst mature myeloid cells represented the greatest proportion of the BM. Most mature cells were either WT or contained the independent *ASXL1* mutation (Figure 3.15c). Of note, the *ASXL1*-mutant CH clone had expanded 4-fold at EOC1 compared to baseline (Figure 3.15b). Whilst leukaemic clones were still present, they were shrinking, and no novel leukaemic clones were selected. At EOC9, 8 months later, blood production was largely sustained by *ASXL1*-mutated and WT hematopoiesis that fully outcompeted leukemic clones within all HSPC compartments (Figure 3.15c). In fact, at EOC9, the A clone contributed to 41% of all HSPCs. Hence, treatment provided a selective advantage for the outgrowth of CH.

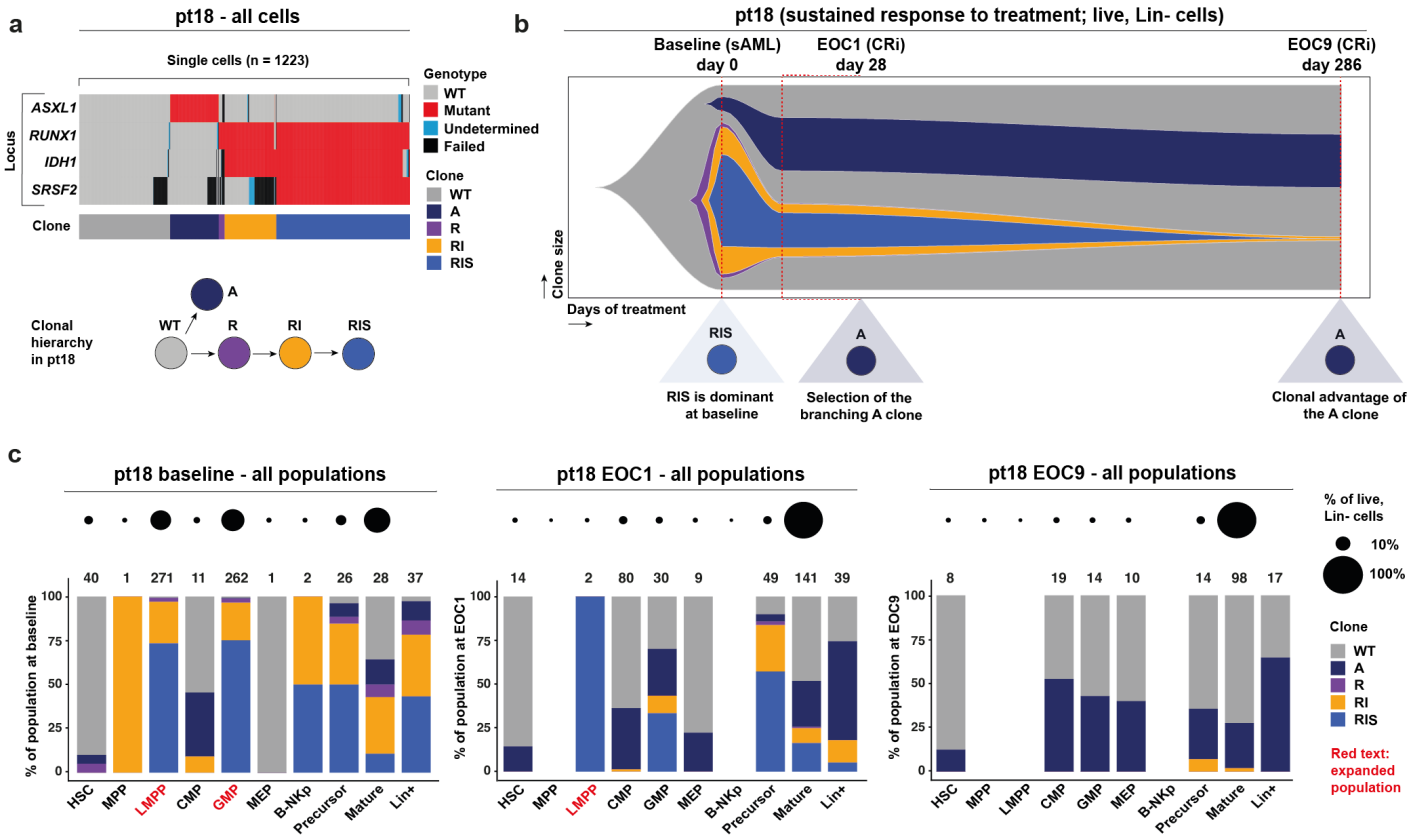


Figure 3.15: Longitudinal clonal analysis for pt18. (a) Same as in Figure 3.8a but for pt18. Somatic mutations as in Figure A.13. Clone abbreviations: WT: wild-type (gray). A: *ASXL1*-mutant (dark blue). R: *RUNX1*-mutant (purple). RI: *RUNX1*/*IDH1*-mutant (orange). RIS: *RUNX1*/*IDH1*/*SRSF2*-mutant (blue). **(b)** Same as in Figure 3.8b but for pt18. **(c)** Same as in Figure 3.8c but for pt18. Clone colour-coding as in (a).

Conversely, the RI clone was present at small frequency exclusively within more downstream GM precursors and mature myeloid cells (Figure 3.15c), consistent with a differentiation effect of therapy on this clone. Whilst *ASXL1* mutations confer unfavorable prognosis in AML [171, 172], persistent *ASXL1* mutations at remission do not correlate with relapse in intensively treated patients [262]. The present data shows that *ASXL1*-mutated CH can expand under ivosidenib/venetoclax/azacitidine and sustain remission, in conjunction with WT cells. Given that *ASXL1* mutations alone do not have transforming potential [281, 377, 378] and that no resistant clones with transforming mutations were selected, pt18 attained durable remission.

In pt10, most of the BM at baseline was composed of GM precursor-like and mature myeloid cells. The expanded GM precursor-like compartment was composed by 93% from a clone harboring 8 recurrent myeloid somatic mutations (genotyping thresholds in Figures A.14a-b), which differentiated very inefficiently (Figures 3.16a-c and Table A.12). In fact, mature myeloid cells were mainly WT. By EOC3 (CRi), WT hematopoiesis was reconstituted, and all leukemic clones were eradicated (Figures 3.16b-c and Table A.2). This was maintained at EOC11, and no alternative leukaemic clones bearing transforming lesions were selected. The patient has remained in remission after 57 months of follow-up. Bulk NGS at remission revealed a *GNAS* mutation at 1% VAF, undetected pre-treatment (Table A.2). Given the small VAF, I did not genotype this locus in TARGET-seq+, and the mutation is likely to be harbored within an independent CH clone.

Thus, CH clones harbouring a single pre-leukaemic driver mutation accompany WT hematopoiesis during sustained remission.

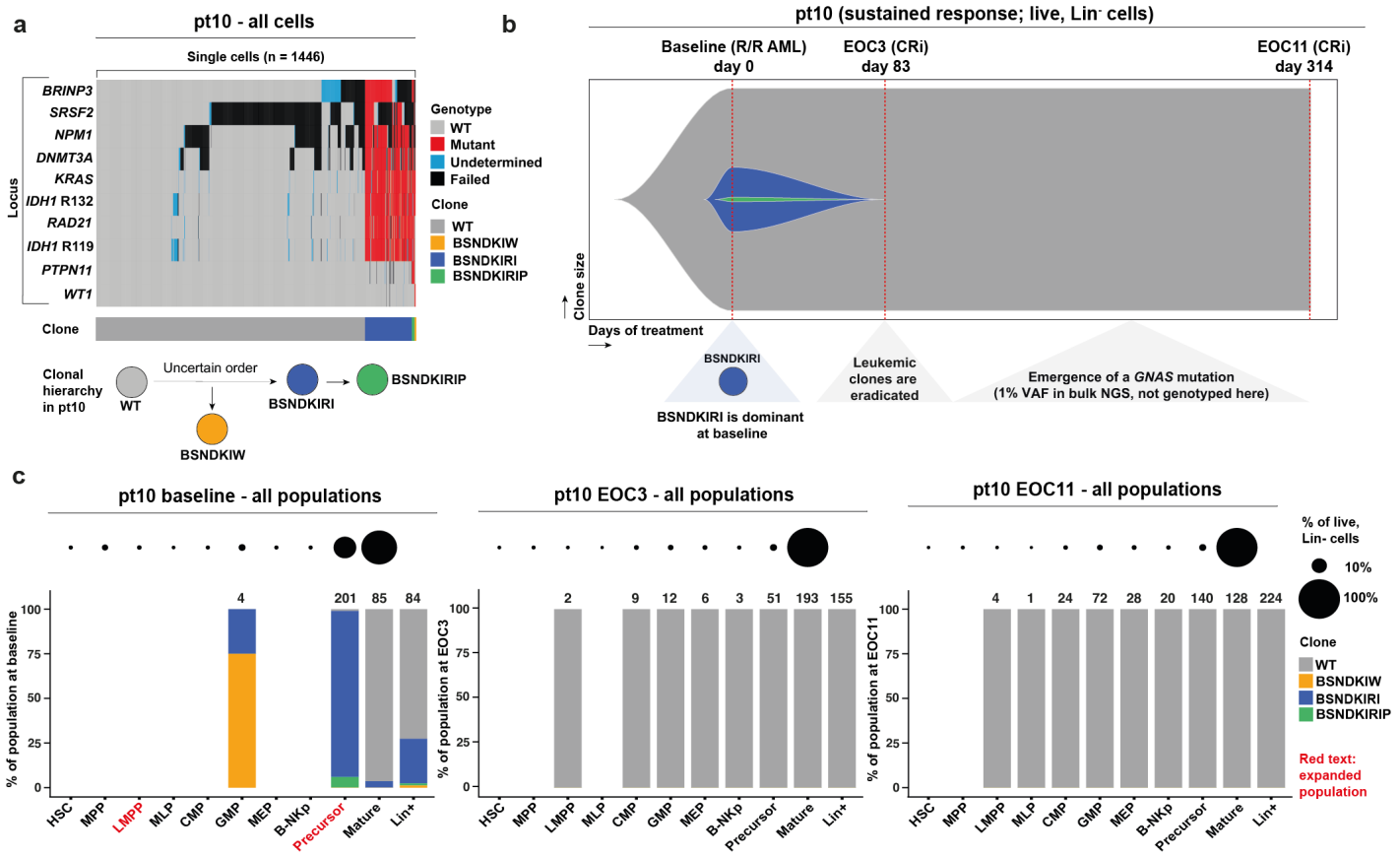


Figure 3.16: Longitudinal clonal analysis for pt10. **(a)** Same as in Figure 3.8a but for pt10. Somatic mutations as in Figure A.14. Clone abbreviations: WT: wild-type (gray). BSNDKIW: *BRINP3*/*SRSF2*/*NPM1*/*DNMT3A*/*KRAS*/*IDH1*-R132S/*WT1*-mutant (orange). BSNDKIRI: *BRINP3*/*SRSF2*/*NPM1*/*DNMT3A*/*KRAS*/*IDH1*-R132S/*RAD21*/*IDH1*-R119P-mutant (blue). BSNDKIRIP: *BRINP3*/*SRSF2*/*NPM1*/*DNMT3A*/*KRAS*/*IDH1*-R132S/*RAD21*/*IDH1*-R119P/*PTPN11*-mutant (green). **(b)** Same as in Figure 3.8b but for pt10. **(c)** Same as in Figure 3.8c but for pt10. Clone colour-coding as in (a).

3.6 Summary of Chapter 3 and proposed model

Taken together, the above data indicate that early clonal selection presages distinct patient outcomes. In patients who relapsed, resistant clones were selected within 3 cycles of treatment. Clonal selection occurred in specific $CD34^+$ progenitor-like or $CD34^-CD117^+$ GM precursor-like BM compartments, which differed between patients (indicating heterogeneous cellular and genetic avenues towards resistance) (Figure 3.17a). Importantly, whilst clonal expansion early during treatment was not always evident by examination of whole BM Lin^- cells (Figures 3.11b, 3.12b, 3.13b), selection was operative specifically within small LSC compartments (Figure 3.17a). Hence, clonal selection within LSCs prior to uncontrolled cellular proliferation is a harbinger of relapse. At the same time, in these cell compartments, there was a reduction of dominant pre-treatment clones, which were sensitive to therapy (Figure 3.17a). In patients experiencing early relapse, the contribution of WT (Figure 3.17a) or CH/pre-leukaemic (Figure 3.17b) cells to progenitor and GM precursor compartments was limited throughout treatment. By contrast, in 4 patients attaining DOR $>$ 2 years (2 of whom remained in sustained remission), there was a higher contribution of WT and/or CH/pre-leukaemic cells throughout treatment. However, whilst in 2/4 of these patients, resistant leukaemic clones ultimately outcompeted WT cells and CH/pre-leukaemic clones, leukaemic clones were virtually eradicated from progenitor and GM precursor compartments in the 2 patients who stayed in durable remission.

The limited cohort size precludes statistically robust associations of genetic aberrations and clinical responses. Nevertheless, patients relapsing with mutations activating signaling pathways (pts 9/11/15) had a significantly shorter OS compared with patients relapsing with cytogenetic abnormalities (pts 4/5/14; Figure 3.18a). This suggests that signaling mutations may provide a more potent proliferative advantage in context of this treatment, warranting careful clinical monitoring.

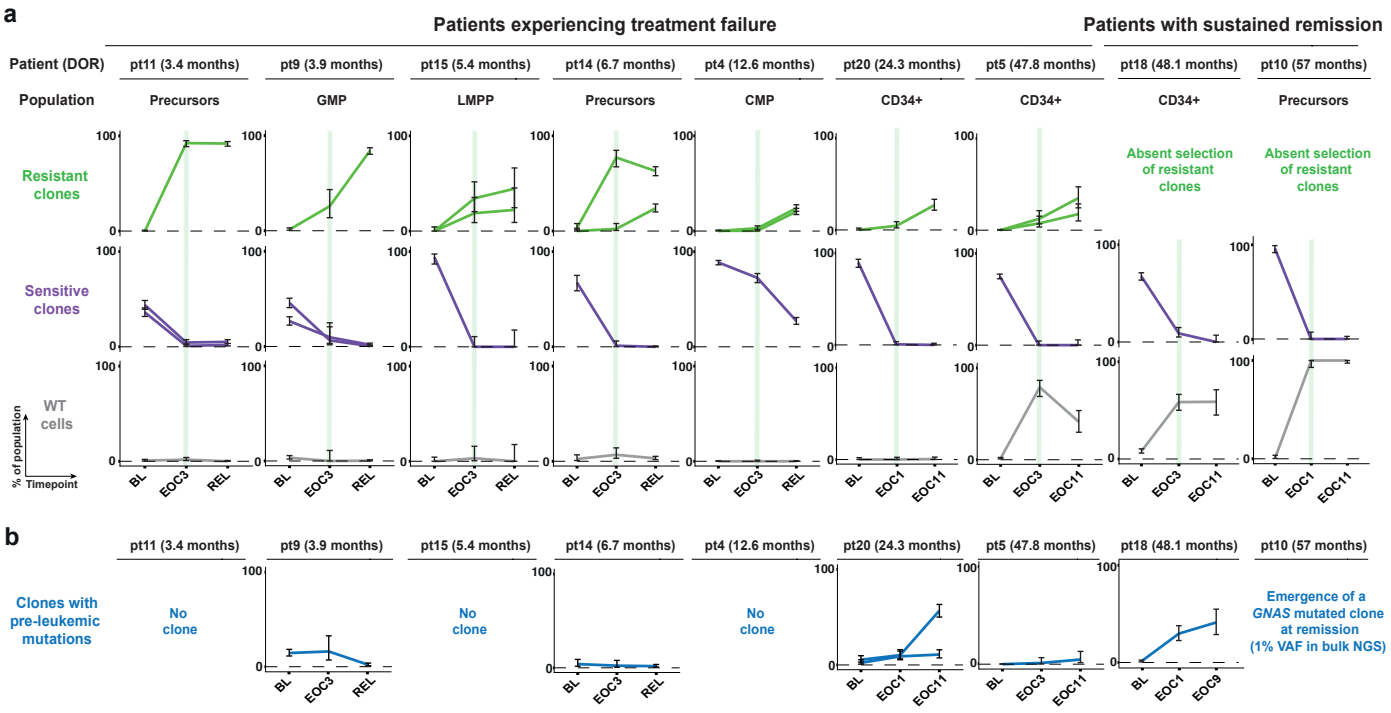


Figure 3.17: Rapid clonal selection within specific immunophenotypic cell populations. **(a)** Longitudinal changes in clonal composition of specific immunophenotypically-defined populations in all studied patients. Clones associated with resistance are indicated in green; dominant pre-treatment clones in purple; WT cells in grey. Error bars depict the 95% CI calculated using the Wilson score interval. For those patients where a relapse sample was available, the most expanded population at relapse was selected. **(b)** Same as in (a) but showing clones harboring only pre-leukaemic mutations (A in pt9; D in pt14; TTAS/TTA in pt20; T in pt5; A in pt18).

Conversely, there was no significant difference between the two groups in terms of the time from early treatment to relapse (Figure 3.18b). This suggests that, whilst both groups of genetic lesions can cause relatively early treatment failure, relapses associated with signaling mutations are more difficult to manage clinically.

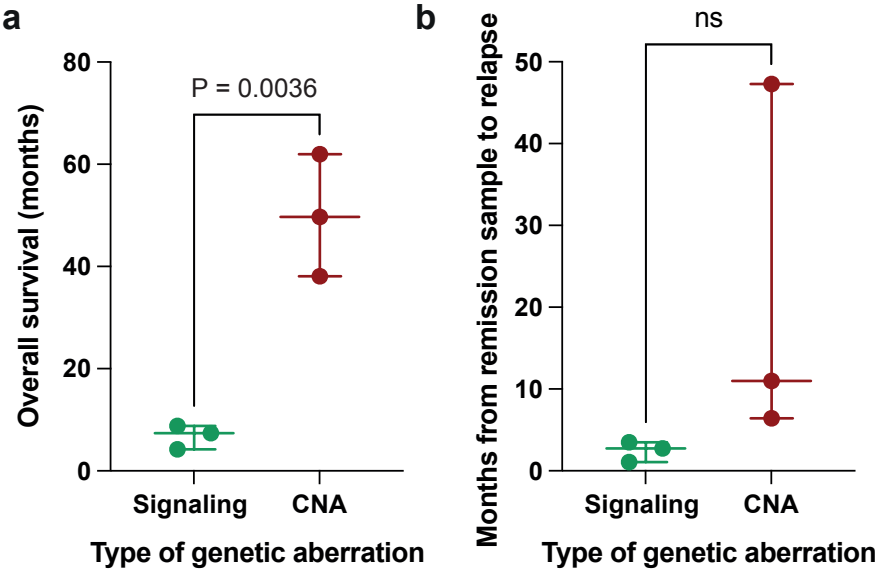


Figure 3.18: Correlation of resistance-associated somatic aberration and response to treatment. (a) OS between patients relapsing with signalling mutations versus patients relapsing with CNAs (brown). Student's t-test was used. **(b)** As in (a) but showing the time (months) between the early treatment time point and treatment failure.

Taken together, I propose a model whereby clonal selection by IVO/VEN ± AZA occurs rapidly, within 3 months of therapy initiation, and presages longer-term patient outcome (Figure 3.19).

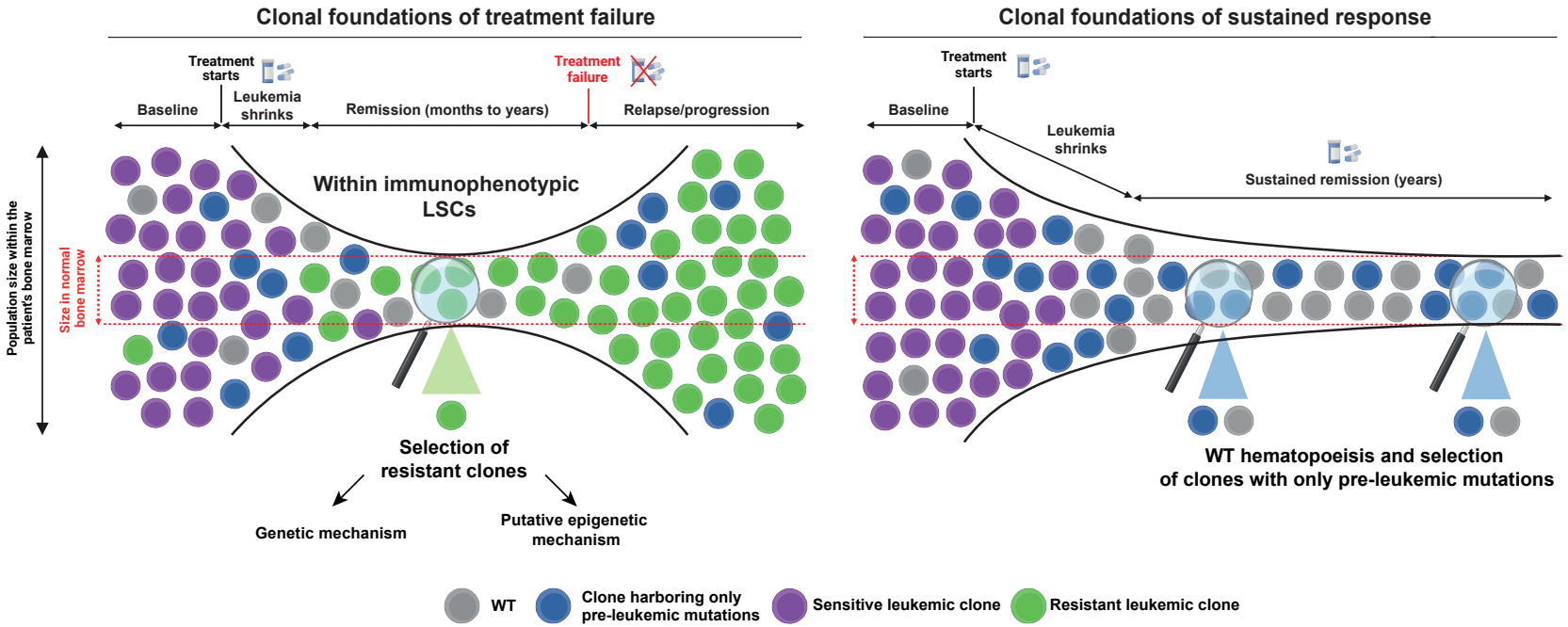


Figure 3.19: Rapid clonal selection within immature BM cells foreshadows patient outcome. Proposed model depicting that clonal foundations of either treatment failure (left) or sustained response (right) are established early during treatment.

Chapter 4

Development and validation of GTAC

Contents

4.1 The rationale for the development of GTAC and overview of the method 97
4.1.1 Overview of GTAC 100
4.2 Development of the GTAC chemistry 102
4.2.1 Integration of targeted genotyping to bulk ATAC 102
4.2.2 Integration of targeted genotyping to scATAC-seq in cell lines 104
4.2.3 A framework for genotyping primer testing prior to execution of GTAC 109
4.3 Validation of GTAC in cell lines 111
4.3.1 Benchmarking GTAC single-cell genotyping in cell lines 111
4.3.2 Allelic drop-out estimation 115
4.3.3 Multiplexed genotyping efficiency in GTAC 117
4.3.4 Benchmarking of scATAC-seq library metrics in GTAC 118
4.3.5 Integration of genotypes and chromatin accessibility profiles in cell lines 124
4.4 Brief summary of Chapter 4 125

I am grateful to Dr. Niels Asger Jakobsen, for insightful discussions and critical advice while developing GTAC. The idea behind GTAC is a joint one. The results presented from section 4.2.3 onwards were published in Cell Stem Cell and STAR Protocols [332, 333] (Appendices A.3A.4)

4.1 The rationale for the development of GTAC and overview of the method

AML evolution is driven by a complex clonal landscape. Rapid selection of therapy-resistant clones occurs within LSC populations, which ultimately propagate relapse. Mechanistic understanding of how distinct genetic clones support LSC phenotypes is thus warranted to tackle resistant disease. Whilst model systems allow to study how mutations corrupt cellular phenotypes (section 1.2.7), these

approaches often fail to reveal the full extent of clonal complexity observed in patients. As exemplified in Chapter 3, clonal and cellular routes of AML evolution are heterogeneous, and co-existing genetic subclones display variable context-specific fitness and adaptability to treatment. In this regard, linkage of single-cell genotyping with gene expression, surface proteome, or DNA methylation, started to disentangle how somatic alterations perturb human molecular landscapes (section 1.2.8). However, our knowledge of how epigenetic regulation is perturbed downstream of mutations is limited.

Chromatin is the medium through which gene expression in regulated in response to extrinsic and intrinsic stimuli. Epigenetic regulation, frequently aberrant in cancer, dictates how and which genes are expressed in which cell types. Epigenetic regulators mediate histone and DNA modifications, which result in downstream changes in chromatin accessibility at cis-regulatory elements. This, in turn, impacts the activity of TFs which drive changes in gene expression. The assay for transposase-accessible chromatin with sequencing (ATAC-seq) is widely adopted to study chromatin accessibility [92, 379, 380]. ATAC-seq uncovered that the landscape of cis-regulatory elements is a strong indicator of BM cell identity [92]. Moreover, recent analyses revealed that some HSC fates, whilst not evident at gene expression level, may be encoded epigenetically, suggesting chromatin defines fate decisions [76]. The advent of single-cell ATAC-seq (scATAC-seq) allowed to study chromatin differences between cell populations within human tissues [88, 380–382], and integration of scATAC-seq and scRNA-seq revealed that variations in chromatin accessibility may precede transcriptional changes [383]. Collectively, these studies suggest chromatin regulates cell potential; it is thus to be expected that LSC potential may be epigenetically encoded in leukaemic cells, downstream of somatic mutations.

scATAC-seq has been linked to genetic lineage tracing by integrated mitochondrial genotyping [384–386]. This allowed to study the differentiation properties of distinct genetic clones and showed that subclones within human malignancies display differential gene regulation. Hence, clonality in both healthy and malignant setting is related to epigenetic variability. More recently, DNA methylation status

at a subset of CpGs whose methylation profile remains stable during differentiation ('epimutations') have also been used to link clonal identity with differentiation landscapes [387]. However, the above methods still do not allow to link coding driver somatic mutations to downstream chromatin variation, given that scATAC-seq libraries capture predominantly regulatory genomic elements, as opposed to coding exonic reads. To overcome this limitation, a method linking targeted single-cell genotyping of exonic genomic loci with chromatin accessibility is required. Such method would (Figure 4.1):

1. Allow to study how single and cooperating somatic mutations corrupt gene regulation in patient samples.

2. Enable to study epigenetic variation within the same genetic clones throughout therapy, revealing non-genetic mechanisms of adaptive resistance.

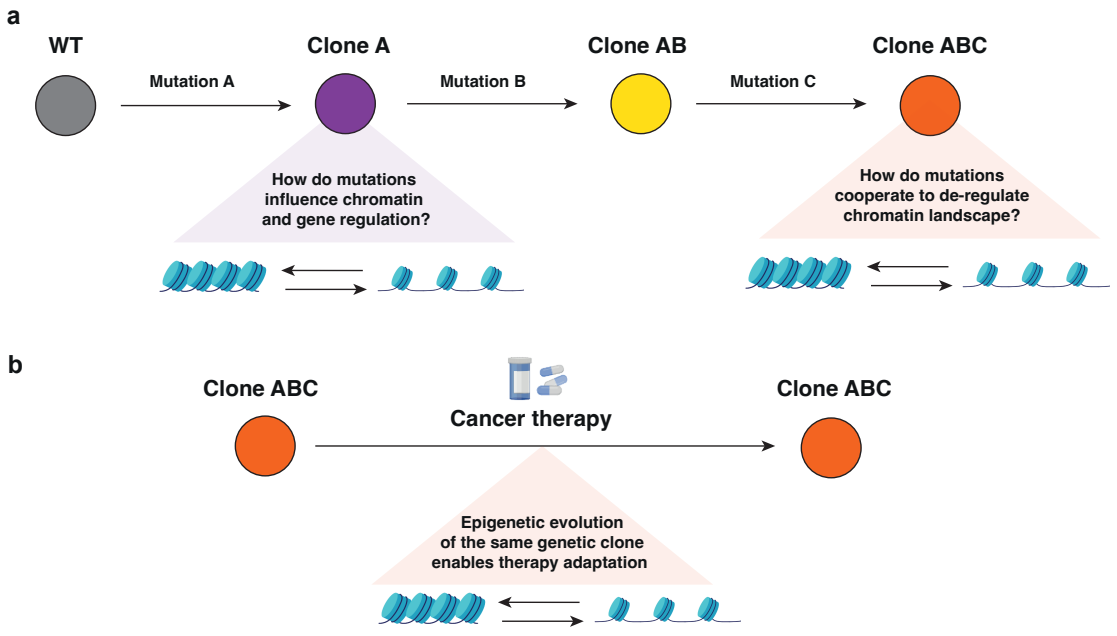


Figure 4.1: Advantages of understanding clone-specific chromatin accessibility. **(a)** A method linking clonal identity and chromatin accessibility would allow to study the consequences of cooperative mutations on gene regulation in primary disease samples. **(b)** This method would also allow to study how the same genetic clone adapts to therapy via epigenetic evolution.

I therefore developed Genotyping with the Assay for Transposase-Accessible Chromatin (GTAC), enabling single-cell genotyping of multiple driver mutations coupled to scATAC-seq [332](Figure 4.2). In this Chapter, a brief overview of GTAC is first provided. I then present the pilot experiments leading to the development of GTAC, followed by quantitative benchmarking in cell lines. Finally, in Chapter 5, I apply GTAC to human AML, in order to study how cooperating mutations in epigenetic regulators and transcription factors alter chromatin landscapes to support leukaemic transformation and LSC maintenance.

4.1.1 Overview of GTAC

A detailed GTAC protocol was published previously [332, 333](Appendices A.3A.4) and is presented in section 2.10. Here, a schematic overview is provided (Figure 4.2). GTAC is based on a plate-based scATAC-seq method developed by Xi Chen and colleagues [349, 350]. A cell population, containing a mixture of genetic clones, is lysed and tagmented by the Tn5 enzyme. Tn5 incorporates i5 and i7 adapter sequences into fragments of open chromatin. Using FACS, single tagmented nuclei are sorted into 384-well plates. A PCR is performed within each well, where uniquely barcoded i5 (plate indexes) and i7 (well indexes) primers anneal to i5 and i7 adapter sequences, amplifying accessible DNA fragments. The main novelty of GTAC consists in leveraging this PCR to simultaneously amplify genomic loci harbouring somatic mutations, using target-specific genotyping primers (non-barcoded; referred to as pre-amplification genotyping primers). Careful pre-amplification genotyping primer design (section 2.10.1) warrants no interference with scATAC-seq fragment amplification. This reaction yields fully indexed scATAC-seq libraries and genotyping amplicons which are not yet indexed. Following the PCR, 80% of the material is pooled to purify sequencing-ready scATAC-seq libraries. The remaining material is transferred to a fresh plate and used for further genotyping target amplification and barcoding. This occurs in two genotyping PCRs (PCR1/2), a strategy adapted from TARGET-seq [80, 329].

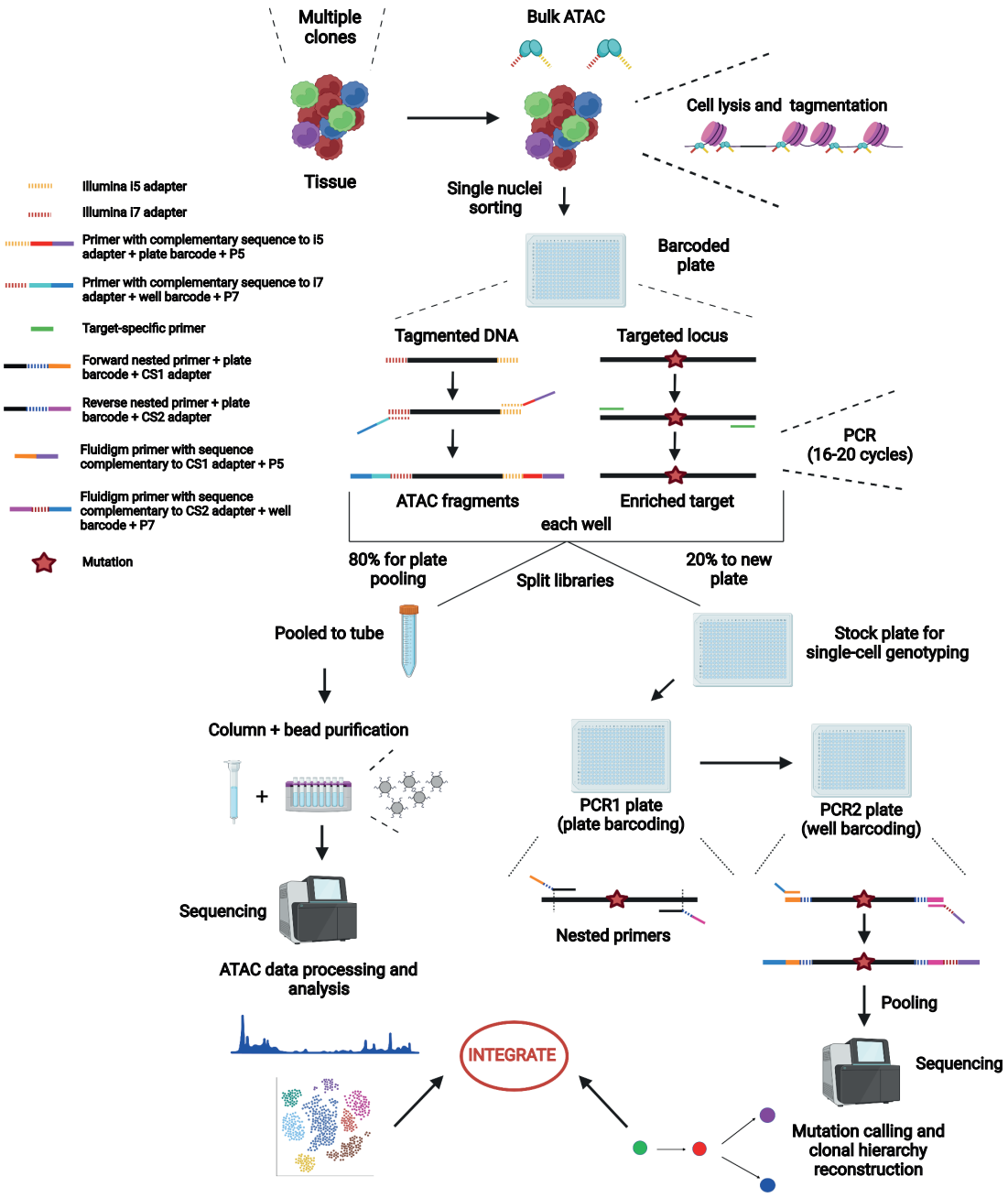


Figure 4.2: Schematic representation of the GTAC method. Primer legend on the left. The mutant locus is indicated by the red star. For technical details, please refer to the main text, section 2.10, and Appendices A.3,A.4. Adapted from [332].

Briefly, in PCR1, nested genotyping primers attach plate barcodes to genotyping amplicons. In PCR2, well barcodes and sequencing adapters are attached to the amplicons generated during PCR1. scATAC-seq and genotyping libraries are sequenced separately. After data pre-processing and single-cell mutation calling at

each locus, clonal information is linked to chromatin accessibility of each single cell, by matching plate and well barcodes from scATAC-seq and single-cell genotyping libraries.

Section 4.2 describes the pilot experiments testing the feasibility of GTAC and the optimal PCR conditions. Section 4.3 presents a quantitative benchmark of scATAC-seq and single-cell genotyping libraries produced by GTAC in cell lines.

4.2 Development of the GTAC chemistry

4.2.1 Integration of targeted genotyping to bulk ATAC

As a starting point, I needed to establish whether targeted genomic locus amplification was compatible with the ATAC chemistry. To test this, I first investigated whether integration of targeted genotyping to standard bulk ATAC would generate a successful genotyping amplicon, whilst preserving the capture of accessible chromatin. I performed two bulk ATAC reactions (using the Omni-ATAC protocol)[348] on ≈50,000 OCI-AML3 cells, each. In one condition, at the PCR step, I used genotyping primers targeting the *DNMT3A*-R882 locus together with standard ATAC primers. The second condition was normal Omni-ATAC without genotyping primers. ATAC libraries showed the expected nucleosomal pattern and library shape was qualitatively comparable between conditions (Figure 4.3a, left). Both conditions subsequently underwent a second PCR, to further amplify the *DNMT3A*-R882 locus for visualization on a 2% agarose gel (in this second reaction, genotyping primers were added to both conditions). The genotyping amplicon of the expected size was detected only when genotyping primers were added to the first PCR (Figure 4.3a, right). Next, I adopted the same experimental strategy, this time simulating in bulk the conditions used in the scATAC-seq protocol [349] (section 2.9). This yielded equivalent results to the experiment above (Figure 4.3b). Overall, genotyping primers are required in the ATAC PCR to generate genotyping amplicons and do not interfere with ATAC library formation.

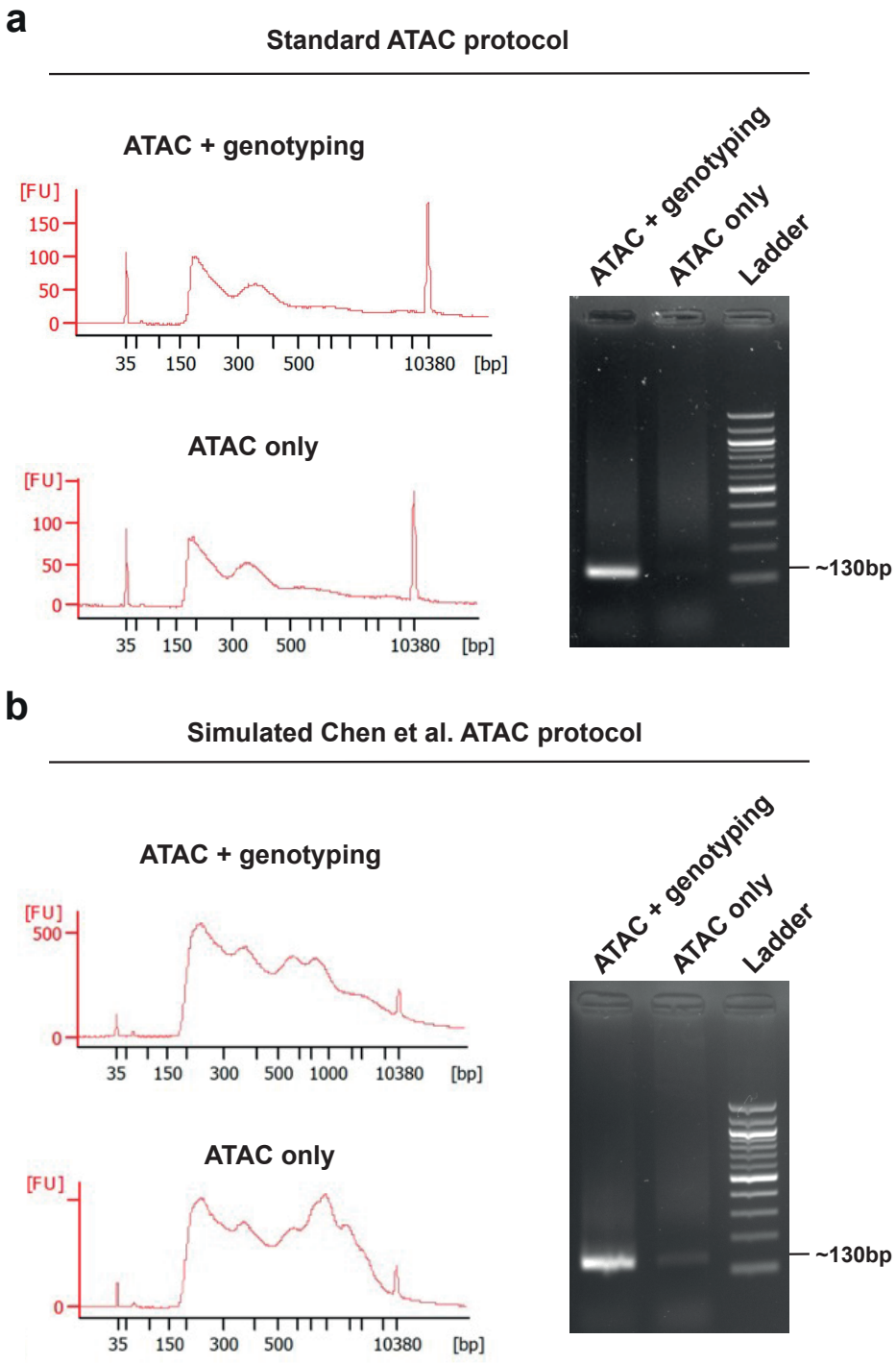


Figure 4.3: Integration of targeted genotyping to bulk ATAC. (a) Left: ATAC libraries derived from 50,000 OCI-AML3 cells each, with (above) or without (below) pre-amplification genotyping primers for the *DNMT3A*-R882 locus. Right: PCR1 genotyping products when pre-amplification genotyping primers were (left) or were not (right) added to the ATAC PCR. The expected amplicon size is 130 bp. **(b)** As in (a) but simulating the single-cell ATAC-seq protocol from Chen et al. [349] in bulk.

4.2.2 Integration of targeted genotyping to scATAC-seq in cell lines

Next, I asked whether targeted genotyping was compatible with scATAC-seq. scATAC-seq can be performed either via droplet-based [381, 388] or plate-based [349, 350] methods. Although droplet-based methods provide higher cell throughput, prior integration of targeted genotyping into droplet-based strategies resulted in limited genotyping efficiency and high allelic drop-out (ADO) rates (a phenomenon in which one allele fails amplification, resulting in mis-genotyping of single cells) [325, 388]. In contrast, I reasoned that, in a plate-based assay, the entire cellular genome in each well is available for genotyping (intact tagmented nuclei are sorted), potentially enabling higher genotyping rates. I thus opted for the plate-based scATAC-seq approach [349, 350]. All pilot experiments in this section were performed using single K562 cells.

I first defined the optimal number of PCR cycles for amplification of accessible chromatin fragments in single K562 cells, via qPCR (Figures 4.4a-c), as suggested previously [349]. This identified 16-18 PCR cycles were optimal for K562 cells (Figure 4.4c). Next, I performed either standard scATAC-seq (termed ATAC only), or scATAC-seq in which pre-amplification *DNMT3A*-R882 genotyping primers were added to the ATAC PCR (Figure 4.5a). To define the optimal genotyping primer concentration in single cells, increasing concentrations (100, 200, 400, or 800 nM) of genotyping primers were used. These tests were performed on two 96-well plates, containing either 1.25 μM or 200 nM of i7/i5 ATAC primers. After PCR, the material from each well was split in two: one part was used to check scATAC-seq libraries while the other was used to simulate PCR1 and check for genotyping amplicon generation. Whilst PCR1 did not yield single-cell genotyping amplicons in the ATAC only condition, successful single-cell genotyping amplicons of the expected size were generated in all conditions where pre-amplification genotyping primers were used (Figure 4.5b). Higher levels of non-specific amplification were observed at increasing genotyping primer concentrations, particularly when i7/i5 ATAC primers were used at 200 nM. Moreover, scATAC-seq library yield was lower when i7/i5 ATAC primers were used at 200 nM (Figure 4.5c). I thus concluded that a condition with ATAC

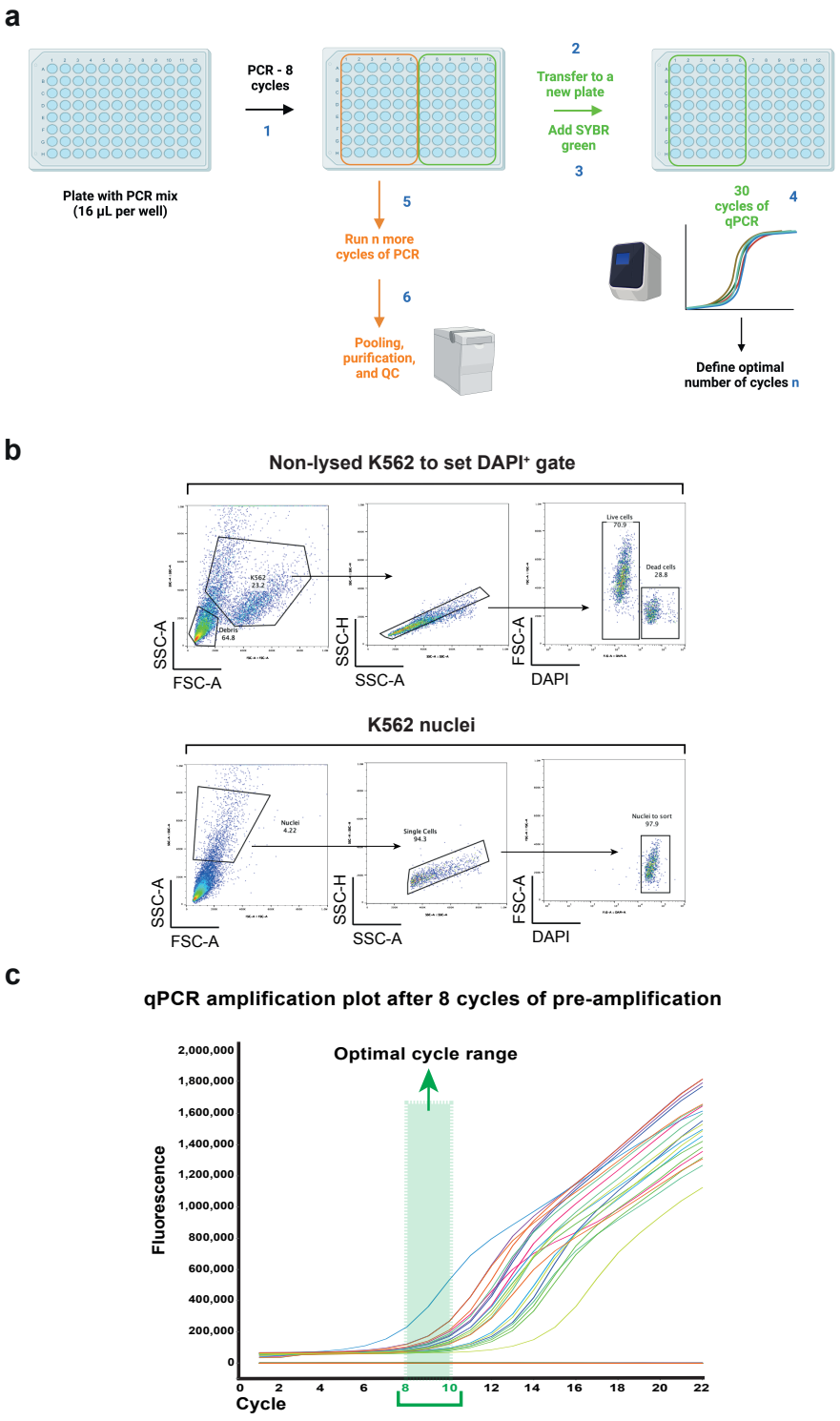


Figure 4.4: PCR cycle number determination in scATAC-seq. **(a)** Strategy for PCR cycle determination in plate-based scATAC-seq. **(b)** FACS strategy for single nucleus sorting. Above: non-lysed K562 cells; Below: tagmented K562 nuclei. **(c)** Representative qPCR amplification plot for scATAC-seq libraries. Each curve represents DNA amplification from a single K562 nucleus. The green shading indicates the optimal number of additional PCR cycles. Panels (a) and (c) adapted from [333].

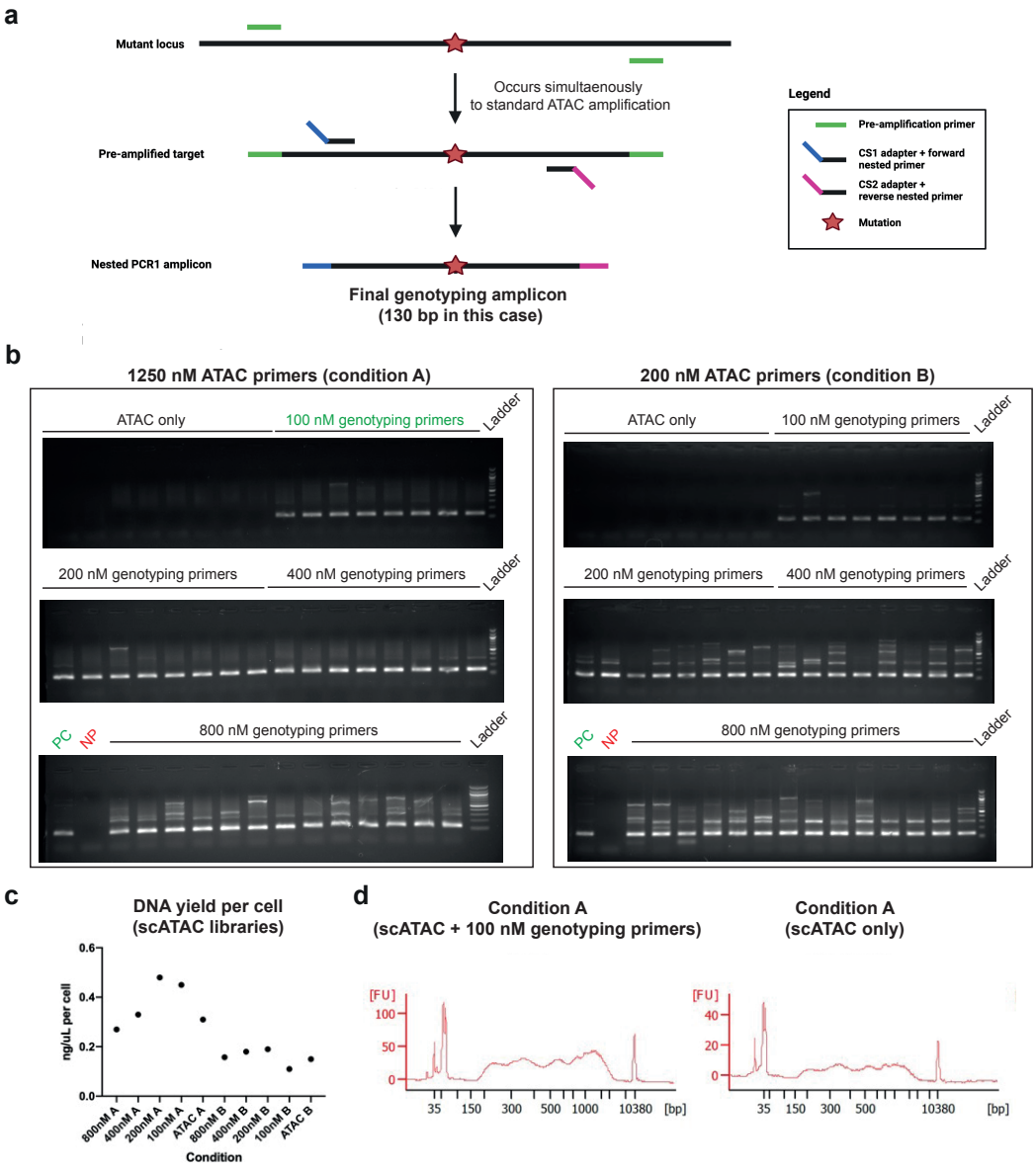


Figure 4.5: Optimization of genotyping primer concentration. **(a)** Schematic of single-cell genotyping in GTAC. Primer legend on the right. The first PCR amplifies both ATAC and genotyping fragments. The second PCR (PCR1) amplifies further the genotyping fragments. Adapted from [333]. **(b)** Single-cell genotyping results when GTAC was performed on K562 cells, using genotyping primers for the *DNMT3A*-R882 locus (130bp). Each lane corresponds to a single sorted nucleus. ATAC primers were used at either 1.25 μM (left, condition A) or 200 nM (right, condition B). For each condition, pre-amplification genotyping primers were added at 100, 200, 400, or 800 nM, or were not added (ATAC only). PC, positive control (gDNA); NC, negative control (empty well). A 100bp ladder was used. **(c)** Average DNA concentration (in ng/μL) per single well for each condition (A or B) from (b). 'ATAC' indicates the condition without genotyping primers. **(d)** Pooled scATAC-seq libraries from 8 single K562 nuclei. Left, 100 nM pre-amplification genotyping primers for the *DNMT3A*-R882 locus were added; right, no genotyping primers were used.

primers at 1.25 μM and genotyping primers at 100 nM was optimal. Importantly, scATAC-seq libraries (derived from a pool of 8 single nuclei) displayed the expected periodic nucleosomal pattern and were qualitatively comparable between conditions with or without genotyping primers (Figure 4.5d). Although DNA yields from single nuclei were too low for high-confidence ATAC library evaluation at single-cell resolution, granular inspection of Bioanalyzer readouts confirmed that ATAC libraries were generated from single nuclei in presence of genotyping primers (Figure 4.6). Collectively, simultaneous scATAC-seq and single-cell targeted genotyping are feasible. I termed this method GTAC.

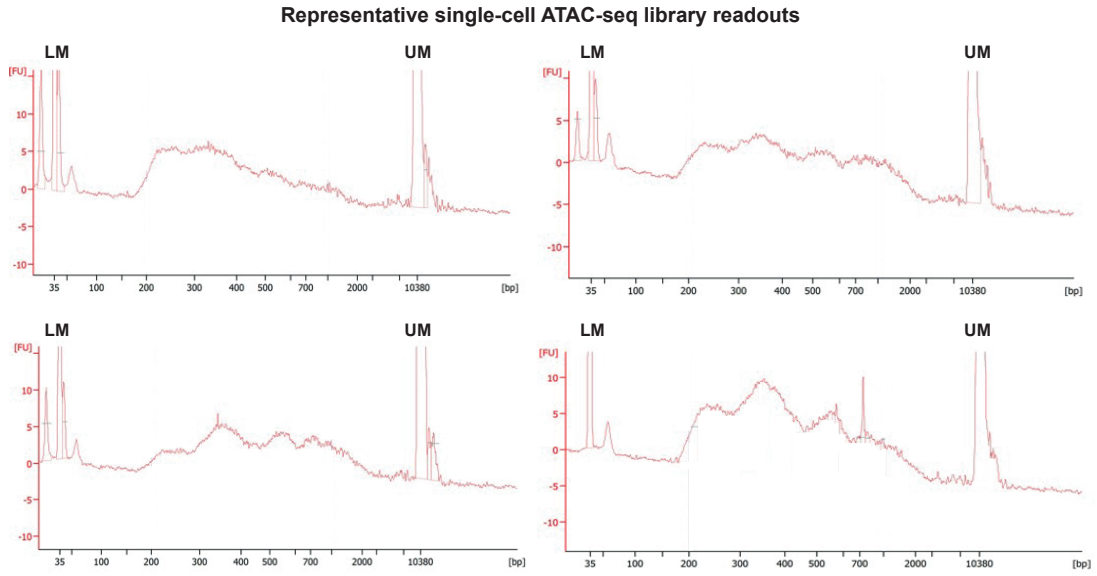


Figure 4.6: GTAC scATAC library traces at single-cell resolution. scATAC-seq libraries generated from single tagmented K562 nuclei. LM, lower marker; UM, upper marker.

In adult cancer, malignant cells frequently contain cooperative mutations acquired in complex clonal hierarchies. Therefore, for GTAC to be applicable to most adult malignancies, multiple genomic loci should be genotyped from the same single cells. I tested GTAC on single K562 nuclei by adding 2 or 3 genotyping primer pairs targeting loci on *DNMT3A*, *TET2*, and *ASXL1* genes. Multiple amplicons of the expected sizes were successfully generated from the same single cells (Figure 4.7a).

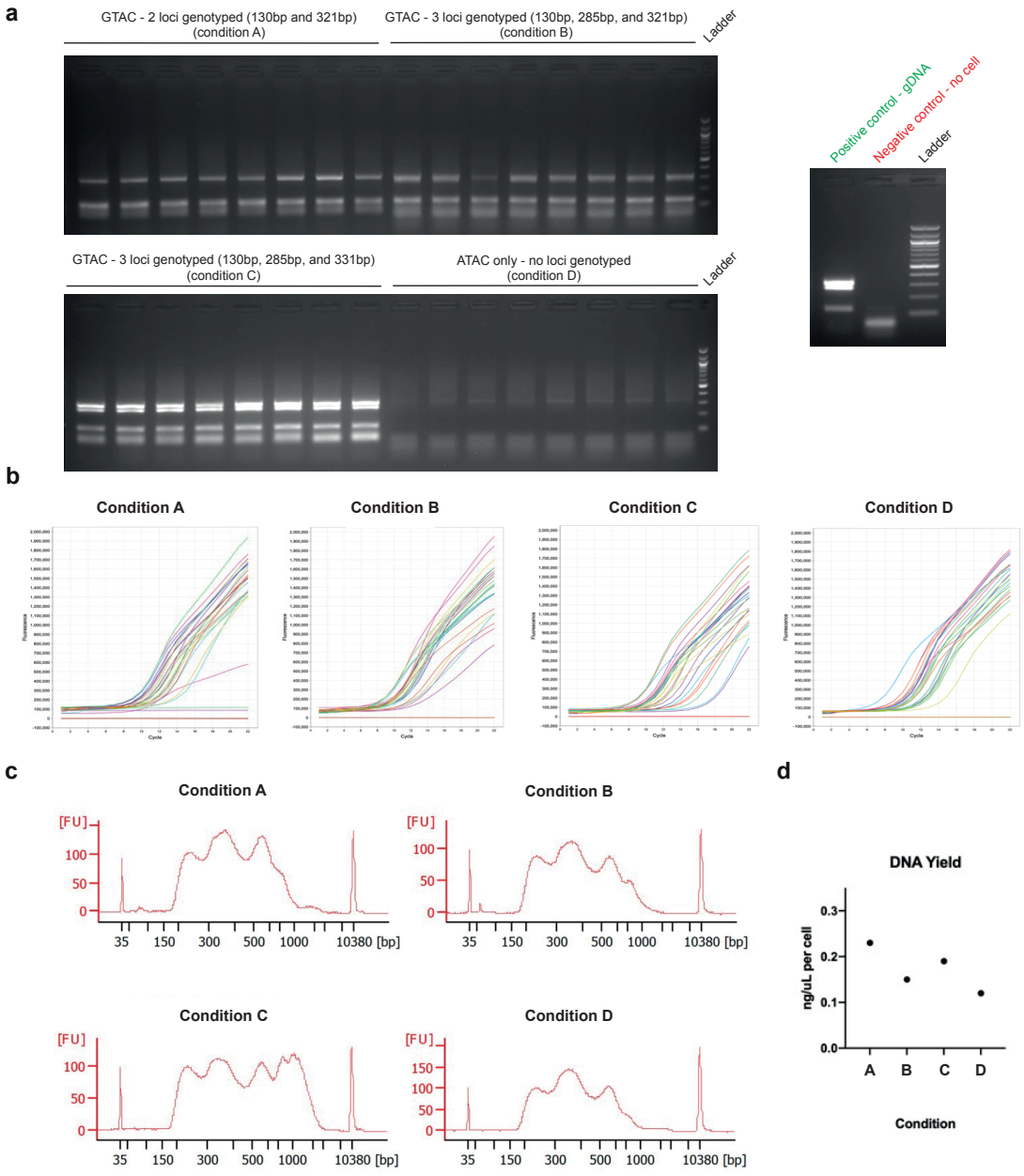


Figure 4.7: Multiplexed single-cell genotyping testing in GTAC. (a) Left, the same experiment as in Figure 4.5b (100 nM genotyping primers and 1.25 μM ATAC primers), but adding genotyping primers for 2 or 3 loci. Each lane corresponds to a single nucleus. The expected amplicon sizes indicated above. In condition D, no genotyping primers were used. Right, gDNA was used as positive control, an empty well as negative control. **(b)** qPCR amplification plots as in Figure 4.4c but relative to the 4 conditions from (a). **(c)** scATAC-seq libraries from pooled K562 single nuclei processed with the 4 conditions from (a). Each trace derives from a pool of 24 single nuclei. **(d)** Average DNA concentration (in ng/μL) per single well for each condition from (a).

Some amplicons displayed stronger bands than others, suggesting differential amplification efficiency. Increasing the number of genotyping primers did not compromise scATAC-seq libraries, as suggested by qPCR amplification plots (Figure 4.7b). Library shapes (Figure 4.7c) and single-cell DNA yields (Figure 4.7d) were also comparable. Hence, multiplexed single-cell genotyping in GTAC is feasible.

4.2.3 A framework for genotyping primer testing prior to execution of GTAC

Efficient GTAC application on primary patient samples requires the design and validation of patient-specific genotyping primers, as mutations vary between patients. This is particularly relevant for LoF mutations, which are more variable between patients and frequently scattered throughout the gene sequence. The design of patient-specific primer panels requires:

1. Prior bulk sequencing of the patient sample to identify driver somatic mutations to be interrogated by GTAC.

2. Appropriate genotyping primer design (2.10.1)(Appendix A.4).

3. Testing of individual primer pairs to check for (a) efficient genotyping amplicon generation; and (b) scATAC-seq library integrity.

4. Testing the combination of all patient-specific genotyping primers to check for the integrity of scATAC-seq libraries when the whole primer panel is used.

In this regard, the above experiments have established an efficient workflow for the testing and optimization of targeted genotyping primers (Figure 4.8a). Briefly, single tagmented nuclei (from a cell line) are sorted into 96-well plates. Each designed primer pair is tested on ≈8 single cells, using ≈8 single cells as ATAC only control. scATAC-seq libraries from the same condition are pooled (keeping conditions separate) and checked using Bioanalyzer, whilst the remaining material is used to simulate a PCR1 reaction to check for efficient genotyping amplicon generation in single cells.

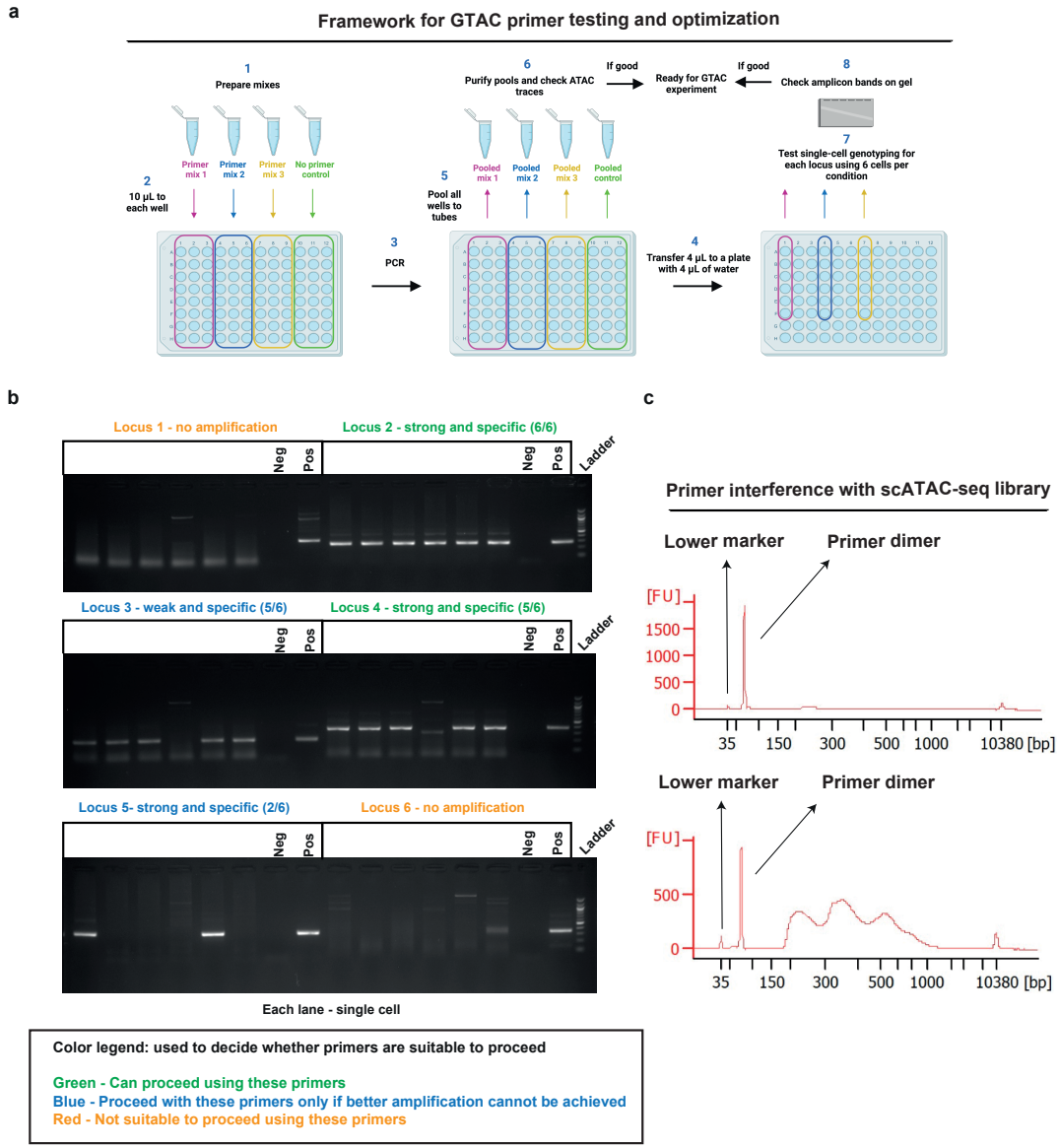


Figure 4.8: Framework for single-cell genotyping primer testing and optimization in GTAC. (a) Schematic of GTAC genotyping primer testing in single cells (Appendix A.4 for details). **(b)** Genotyping primer testing, displaying optimal (green), sub-optimal (blue), or failed (orange) genotyping. Each lane corresponds to a single cell. gDNA used as positive control (Pos); an empty well used as negative control (Neg). **(c)** Sub-optimal scATAC-seq libraries, due to interference of genotyping and ATAC primers.

Primer pairs which do not generate strong and specific bands should be re-designed and re-tested (Figure 4.8b). Furthermore, primer pairs which interfere with i7/i5 ATAC primers (Figure 4.8c) should also be re-designed. This procedure is performed iteratively for each genotyping locus. When primer pairs for all desired loci work well individually, a final test is performed in which all primer pairs are added simultaneously to the ATAC PCR.

4.3 Validation of GTAC in cell lines

To rigorously benchmark both scATAC-seq and single-cell genotyping library quality in GTAC, a series of experiments was performed on several cell lines [332], to answer two main questions:

1. Does GTAC allow for accurate, high-confidence, single-cell genotyping?
2. Does targeted genotyping compromise scATAC-seq library quality?

4.3.1 Benchmarking GTAC single-cell genotyping in cell lines

I first wanted to confirm whether genomic targets were properly captured by single-cell genotyping libraries. I performed GTAC on single K562 cells, with genotyping primers against 5 mutant/variant loci (Figure 4.9a and Table 4.1; *TET2* c.5162T>G is considered variant)[389]. As control, I performed standard scATAC-seq with no genotyping primers.

Gene	Exon	DNA_change	Protein_change	Chr	Cell line	Zygosity
ASXL1	13	c.1773C>A	p.Y591*	20	K562	het
HOXA9	1	c.574G>A	p.D192N	7	K562	het
CBL	15	c.2374G>C	p.D792H	11	K562	het
NOTCH1	6	c.977G>T	p.G326V	9	K562	het
TET2	11	c.5162T>G	p.L1721W	4	K562	het
RUNX1	4	c.445G>A	p.A149T	21	JURKAT	het
NRAS	3	c.182A>T	p.Q61L	1	OCI-AML3	hom
DNMT3A	23	c.2644C>T	p.R882C	2	OCI-AML3	het
NPM1	11	c.860_863dup	p.W288Cfs*12	5	OCI-AML3	het
FLT3	14	c.1775_1795dup	p.E598_y599ins	13	MOLM13	het

Table 4.1: Genomic loci screened by GTAC in cell lines. DNA and aminoacidic change are reported. Het, heterozygous; hom, homozygous.

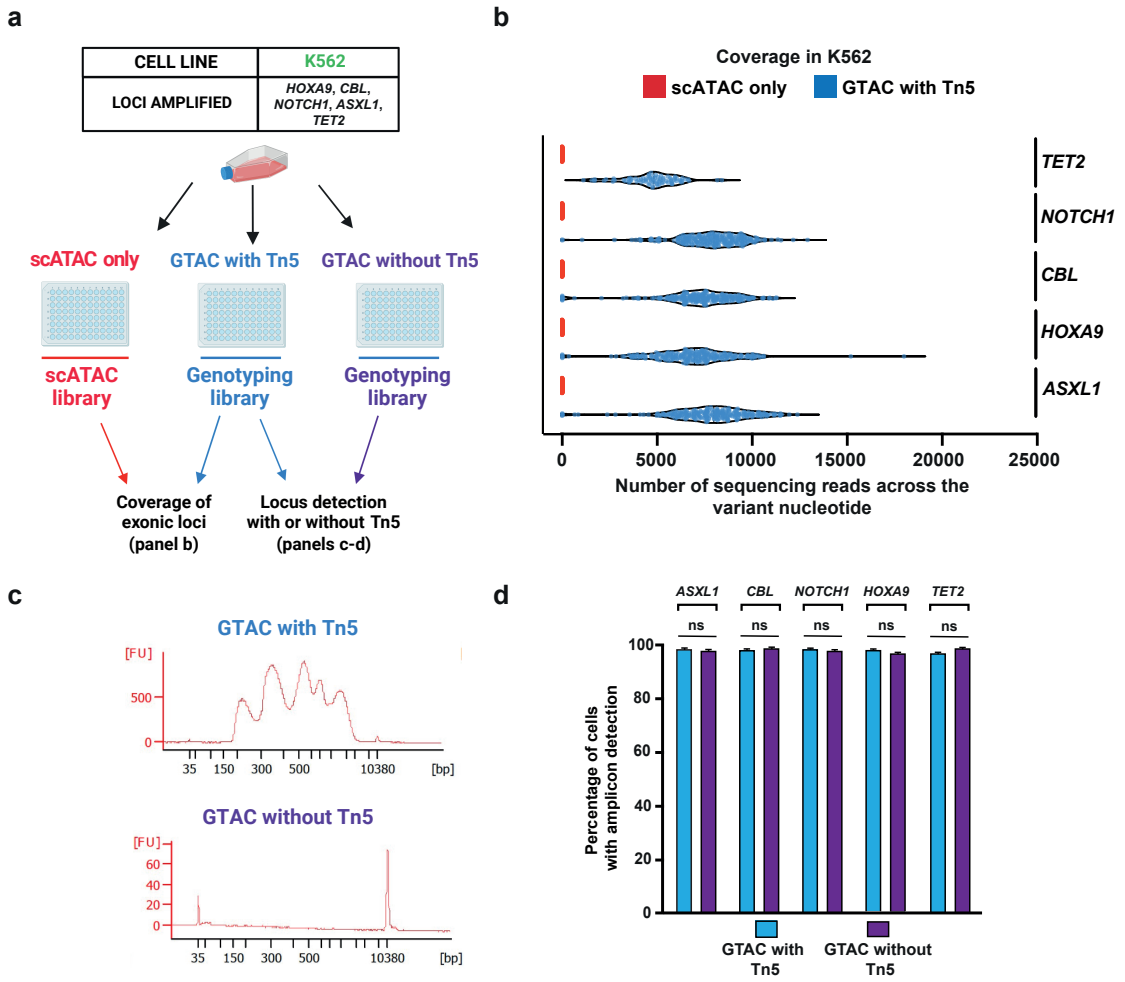


Figure 4.9: Genotyping locus capture in GTAC. (a) Validation of efficient amplicon capture in GTAC single-cell genotyping, in K562. Top, the genotyped loci; below, K562 cells were processed using standard scATAC-seq (red) or GTAC in presence (blue) or absence (purple) of Tn5. **(b)** Sequencing reads (x-axis) mapping to the genotyped loci (right) in single K562 cells (dots), comparing standard scATAC-seq (red) and GTAC (blue) libraries. Median coverage for each locus, for standard scATAC-seq/GTAC: *TET2*: 0/4,823; *NOTCH1*: 0/7,993; *CBL*: 0/7,461; *HOXA9*: 0/6,961; *ASXL1*: 0/7,980. 240 cells were processed by GTAC for *ASXL1*/*CBL*/*HOXA9*/*NOTCH1*; 80 for *TET2*. 95 cells were processed by standard scATAC-seq at each locus. **(c)** Pooled scATAC-seq traces from a GTAC experiment on 190 single K562 cells with (above) or without (below) Tn5. **(d)** Percentage of K562 cells showing $\geq$ 30 sequencing reads mapping to genotyping loci when the Tn5 was (blue) or was not (purple) added. Standard errors are shown. The Fisher's exact test was used.

I compared coverage across the five targeted loci between GTAC and standard scATAC-seq libraries. Whilst all loci were efficiently captured by GTAC in most single cells, the median number of reads mapping to the mutant/variant nucleotide was 0 for scATAC-seq libraries (Figure 4.9b). Expectedly, this confirmed genotyping

primers are required to capture these mutant/variant loci, which are not efficiently covered by standard scATAC-seq libraries. Of note, capture of genomic loci from standard scATAC-seq libraries may be more efficient if the locus is accessible, but genotyping primers bypass this requirement.

Next, I tested whether genotyping amplicon capture is facilitated by Tn5 integration in proximity of the genotyped locus. If Tn5 tagments DNA even hundreds of bp away from the locus, initial amplification of the fragment by i5/i7 primers may facilitate the capture of the genotyping locus. I thus performed GTAC with or without the addition of Tn5 (Figure 4.9a) and calculated the rate of efficient capture of the genotyping locus. Expectedly, the scATAC-seq library was generated only when Tn5 was added (Figure 4.9c). Importantly, genotyping target amplification at all 5 loci was equally efficient with or without Tn5 (Figure 4.9d). This shows that genotyping locus capture does not depend on Tn5 integration, thus being unrelated to chromatin accessibility in proximity of the locus.

To answer whether single-cell genotyping in GTAC was accurate, I analysed the genotype of single K562 cells at the screened mutant/variant loci and, additionally, genotyped 3 other cell lines (MOLM-13, JURKAT, and OCI-AML3) at 5 previously reported mutant loci (Figure 4.10a and Table 4.1). Of note, single-cell genotyping approaches suffer from an error rate, derived from (a) uneven amplification of different alleles from single cells and (b) errors introduced by the DNA polymerase during library preparation and sequencing [80, 332]. These artefacts lead to background noise (i.e. some mutant reads are observed in cells which are WT; section 2.8.2). To control for noise, I genotyped all loci in single THP1 cells as well, WT at all the above loci (Figure 4.10a and Table 4.1). Encouragingly, all five mutant/variant loci in K562 displayed single-cell variant allele frequency (scVAFs) $< 1\%$ in THP1 cells (Figure 4.10b), indicating low degrees of noise. ScVAF distributions in K562 were consistent with all loci being heterozygous and K562 cells being hypotriploid [389] (Figure 4.10b and Table 4.1; in triploid cells, a heterozygous mutation is present on either 1/3 or 2/3 alleles). ScVAF distributions suggested *ASXL1*, *CBL*, *NOTCH1*, and *HOXA9* mutations were present on 1/3 alleles, whilst

the *TET2* variant was present on 2/3 alleles (Figure 4.10b). I also genotyped the heterozygous *RUNX1* locus in JURKAT, and the heterozygous *DNMT3A* and homozygous *NRAS* loci in OCI-AML3 cells [389] (Figure 4.10c; JURKAT and OCI-AML3 are largely diploid).

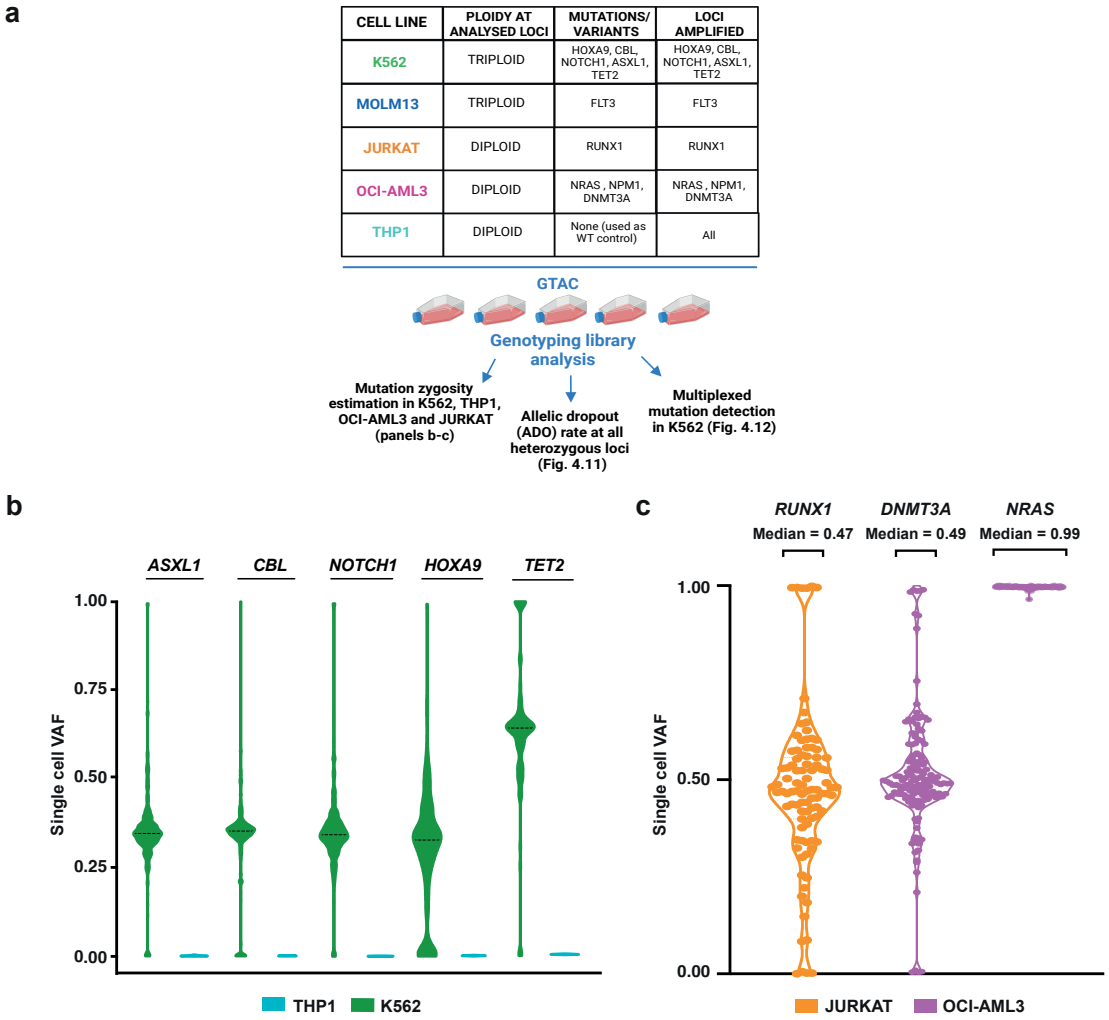


Figure 4.10: Mutation zygosity estimation in GTAC single-cell genotyping. **(a)** Validation of single-cell genotyping accuracy in 5 cell lines. Cell lines, their ploidy at genotyped loci, mutations/variants expected, and genotyping loci amplified are shown. Mutations/variants investigated are listed in Table 4.1. **(b)** scVAF for five mutant/variant loci in K562 cells. Cells with coverage $\geq$ 30 were used. Median scVAFs (dashed lines) in K562/THP1 (THP1 used as WT control): *ASXL1*: 0.345/0.001; *CBL*: 0.352/0.002; *NOTCH1*: 0.342/0.001; *HOXA9*: 0.327/0.002; *TET2*: 0.642/0.001. **(c)** scVAF at the *RUNX1* locus in JURKAT (orange; n = 95 cells) and for *DNMT3A* (purple; n = 151 cells) and *NRAS* (purple; n = 95 cells) loci in OCI-AML3. Each dot is a single cell. Median scVAFs are 0.47, 0.49, and 0.99, for *RUNX1*, *DNMT3A*, and *NRAS*, respectively.

ScVAFs of *RUNX1* and *DNMT3A* mutations suggested a heterozygous state (median

scVAF of 0.47 and 0.49, respectively). In contrast, *NRAS* scVAFs clustered around 0.99 and 1 (median 0.99), suggesting a homozygous state. Overall, single-cell genotyping in GTAC is necessary to capture mutant loci. Mutation calling displays low levels of noise, provides accurate genotypes, and discriminates between heterozygous and homozygous mutant cells.

4.3.2 Allelic drop-out estimation

Allelic drop-out (ADO) is an important technical artifact in single-cell genotyping, with considerable implications for comparisons between genotypes. ADO occurs when one allele is not amplified at single-cell level. If the mutant allele is not amplified in a heterozygous mutant cell, the cell is wrongly assigned as WT. If ADO rates are high, this leads to incorrect genotyping of a substantial proportion of cells, diluting true differences between genotypes. Droplet-based approaches for joint single-cell genotyping and scRNA-seq or scATAC-seq suffer from high ADO rates [325, 388]. In contrast, plate-based methods for joint scRNA-seq and single-cell genotyping, which rely on gDNA for genotyping, have lower ADO rates [80, 329].

Given that all cell lines genotyped above should harbour the mutation/variant within every cell (i.e. there is a ground truth about their genotype), this allowed to benchmark ADO rates in GTAC at all genotyped heterozygous mutant/variant loci. To establish thresholds for mutation/variant calling, I used scVAF distributions and the mutant read number from THP1 WT control cells (Figure 4.11a; section 2.11.2). Briefly, three thresholds were set: (a) the lower scVAF threshold; (b) the upper scVAF threshold; and (c) the minimum number of mutant/variant reads. Cells with scVAF between the two thresholds, and displaying sufficient mutant/variant reads, are called heterozygous mutant/variant (i.e. no ADO had occurred). By contrast, mutant ADO occurred in those cells below the lower scVAF threshold and/or below the minimum number of mutant reads, whereas WT ADO occurred in those cells above the upper scVAF threshold. At all heterozygous mutant/variant loci, ADO rates were low and most cells were called heterozygous mutant/variant (Figure 4.11b).

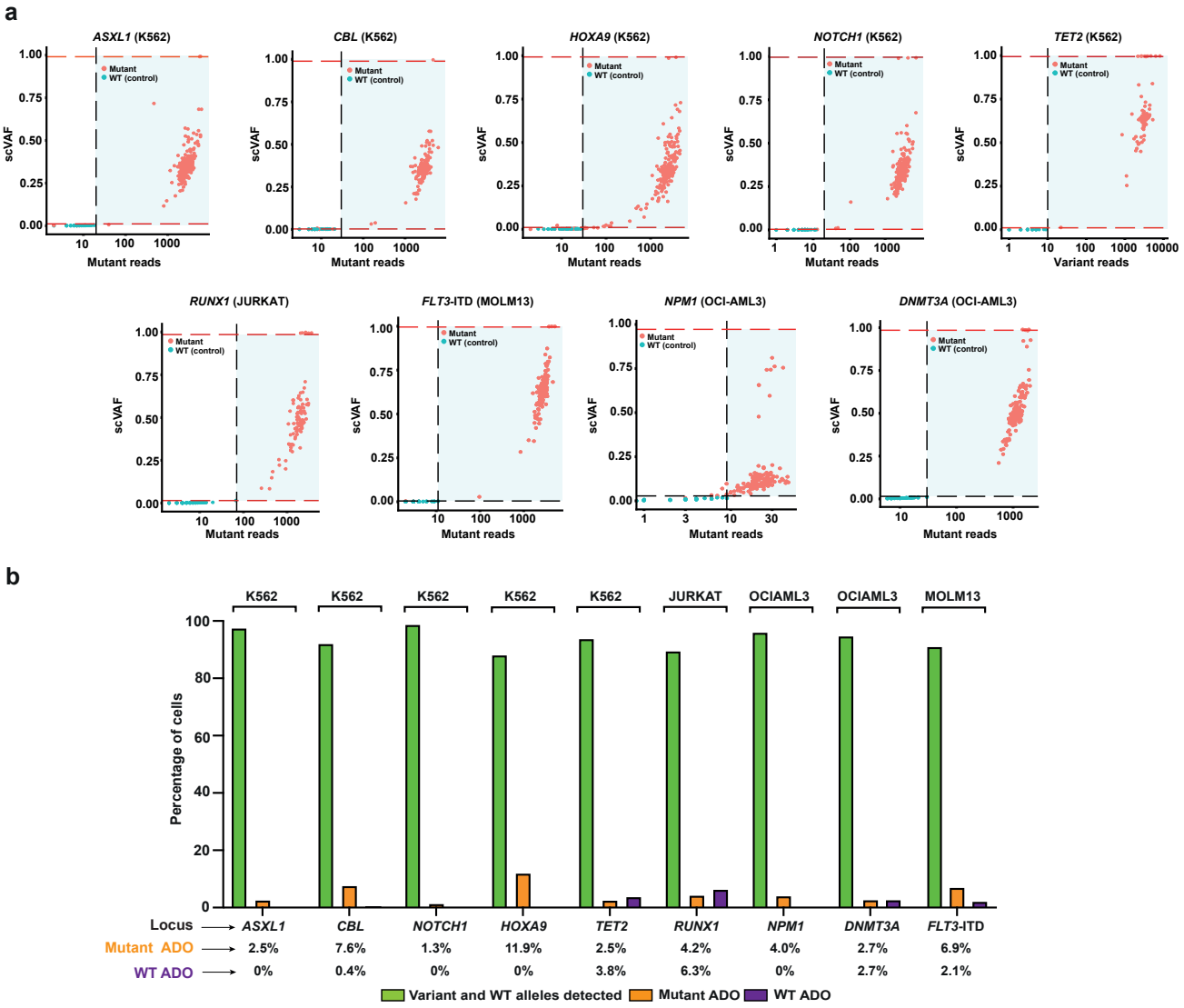


Figure 4.11: *(Legend on next page)*

Figure 4.11 *(previous page)*: ADO estimation in GTAC. (a) scVAF (y-axis) versus the number of mutant/variant reads (x-axis) for heterozygous mutant/variant loci in the respective cell lines (indicated above). Dots are single cells. THP1 cells (cyan) were used as WT control. Dashed lines indicate genotyping thresholds (sections 2.11.1 and 2.11.2). **(b)** ADO rates for the 9 heterozygous loci (x-axis) in the respective cell lines. Green bars indicate the percentage of cells in which both mutant and WT alleles were detected (i.e. no ADO occurred). Purple and orange bars indicate the percentage WT and mutant ADO, respectively. Cell numbers: 236 (*ASXL1*), 238 (*CBL*), 235 (*HOXA9*), 232 (*NOTCH1*), 80 (*TET2*), 93 (*RUNX1*), 149 (*NPM1*), 151 (*DNMT3A*), and 144 (*FLT3*). Only cells passing genotyping QC were used.

Although some variability was observed between loci, ADO rates were low and ranged from 1.3% for *NOTCH1* up to 11.9% for *HOXA9* (Figure 4.11b). Of note, as *ASXL1*, *CBL*, *NOTCH1*, and *HOXA9* loci were found on 1/3 alleles (i.e. 1 mutant allele competes with 2 WT alleles in the PCR), mutant ADO rates were expectedly higher than WT ADO rates. ADO rates were low at loci harbouring *DNMT3A*-R882, *NPM1*-W288, and *FLT3*-ITD mutations, the most frequent hotspot mutations in AML [171, 390](Figure 4.11b). ADO rates in GTAC were comparable to those reported for TARGET-seq+ [80]. GTAC thus enables high-confidence single-cell genotyping.

4.3.3 Multiplexed genotyping efficiency in GTAC

To assign single cells to clones within complex clonal hierarchies, accurate genotyping of multiple loci within the same cells is required. To test whether GTAC allowed for parallel detection of multiple mutations in the same cells, I simultaneously genotyped 5 mutant/variant loci (Table 4.1) in single K562 cells. Simultaneous single-cell detection of all 5 mutations was observed in 79% of cells (Figure 4.12). Encouragingly, at least 4 mutations were detected in 94% of cells. By contrast, only 6% of K562 cells failed single-cell genotyping at 2 or more loci (Figure 4.12). Of note, failed single-cell genotyping of a somatic mutation which occurred early during clonal evolution can still result in a correct clonal assignment if later somatic events are efficiently detected within the same cell (section 2.8.6). This, coupled to the rates of efficient multiplexed single-cell mutation detection (Figure 4.12), suggests GTAC can lead to correct clonal assignments for most single cells. Taken

together, GTAC enables accurate, high-confidence genotyping of multiple genomic loci within single cells.

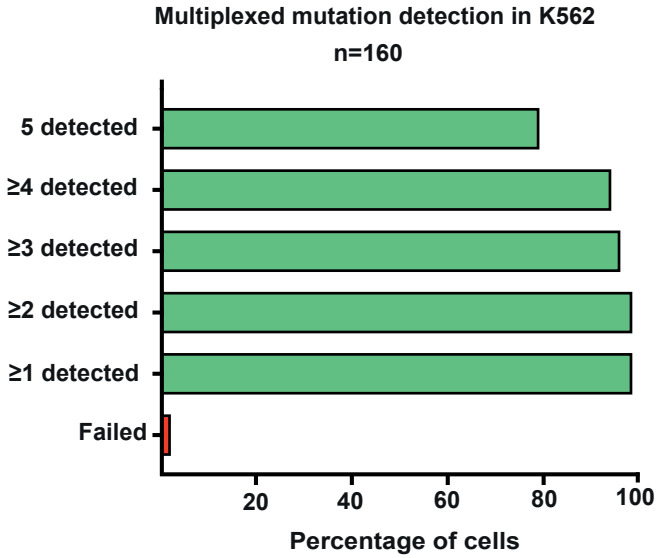


Figure 4.12: Multiplexed genotyping success rates in K562. Frequency of simultaneous detection of multiple mutations in 160 single K562 cells. In total, 5 loci were co-amplified. The percentage of cells in which all 5 (5 detected; 79%), or at least 4 (94%), 3 (96%), 2 (98%) or 1 (98%) mutations were detected is shown. The red bar shows the percentage of cells with failed genotyping at all loci (2%).

4.3.4 Benchmarking of scATAC-seq library metrics in GTAC

To yield high-quality, reliable readouts of chromatin accessibility, scATAC-seq libraries in GTAC should not be impacted by genotyping primers. To test whether the quality of scATAC-seq libraries was perturbed by genotyping, I performed GTAC on K562 cells by adding primers for 4, 6, or 8 genomic loci. I compared scATAC-seq metrics with those from standard in-house plate-based scATAC-seq with no genotyping primers (Figure 4.13a; 80 cells were used for each condition). Metrics were additionally compared with published scATAC-seq K562 data obtained either with a plate-based [349] or 10X Chromium droplet-based approach [391]. Encouragingly, scATAC-seq library traces were qualitatively comparable (Figure 4.13b) and generation of genotyping amplicons was successful in all GTAC conditions, despite the increasing number of genotyped loci (Figure 4.13c).

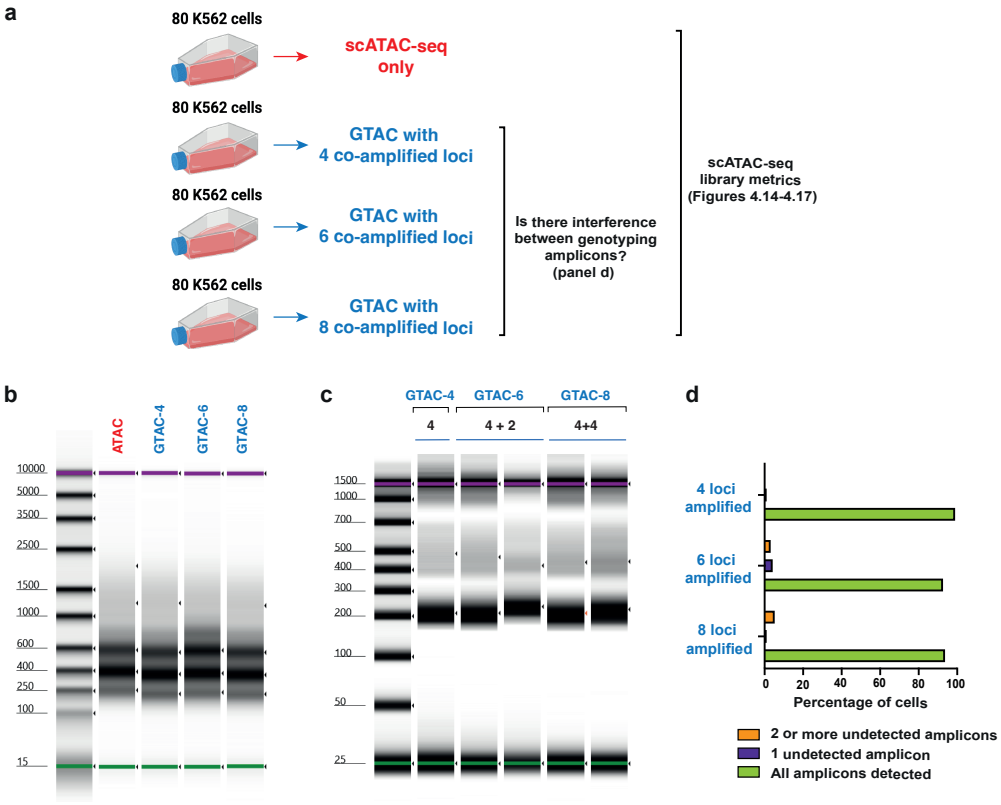


Figure 4.13: GTAC with increasing numbers of genotyped loci. (a) Standard scATAC-seq (red), or GTAC with primers for 4, 6, or 8 genotyping loci, were performed on 80 single K562 cells, each. **(b)** Pooled scATAC-seq libraries from each condition in (a). **(c)** Genotyping amplicons for each GTAC condition in (a). When more than 4 loci were amplified, the genotyping PCR1 reaction was split into two mixes (amplicon number per mix indicated above). **(d)** Percentage of K562 cells displaying successful detection of genotyping loci ($\geq$ 30 sequencing reads), when 4 (top), 6 (middle), or 8 (bottom) loci were co-amplified. Colours show the number of detected amplicons.

Consistently, all genotyping amplicons were detected in $>$ 90% of cells in all GTAC conditions, indicating no substantial interference between genotyping amplicons (Figure 4.13d).

The average number of unique nuclear fragments per single cell (i.e. scATAC-seq library complexity) is one useful metric to assess scATAC-seq library quality. Library complexity was high across in-house GTAC and standard scATAC-seq conditions and there were no significant differences (Figure 4.14a). Moreover, there was a significant improvement in library complexity over previously published scATAC-seq approaches.

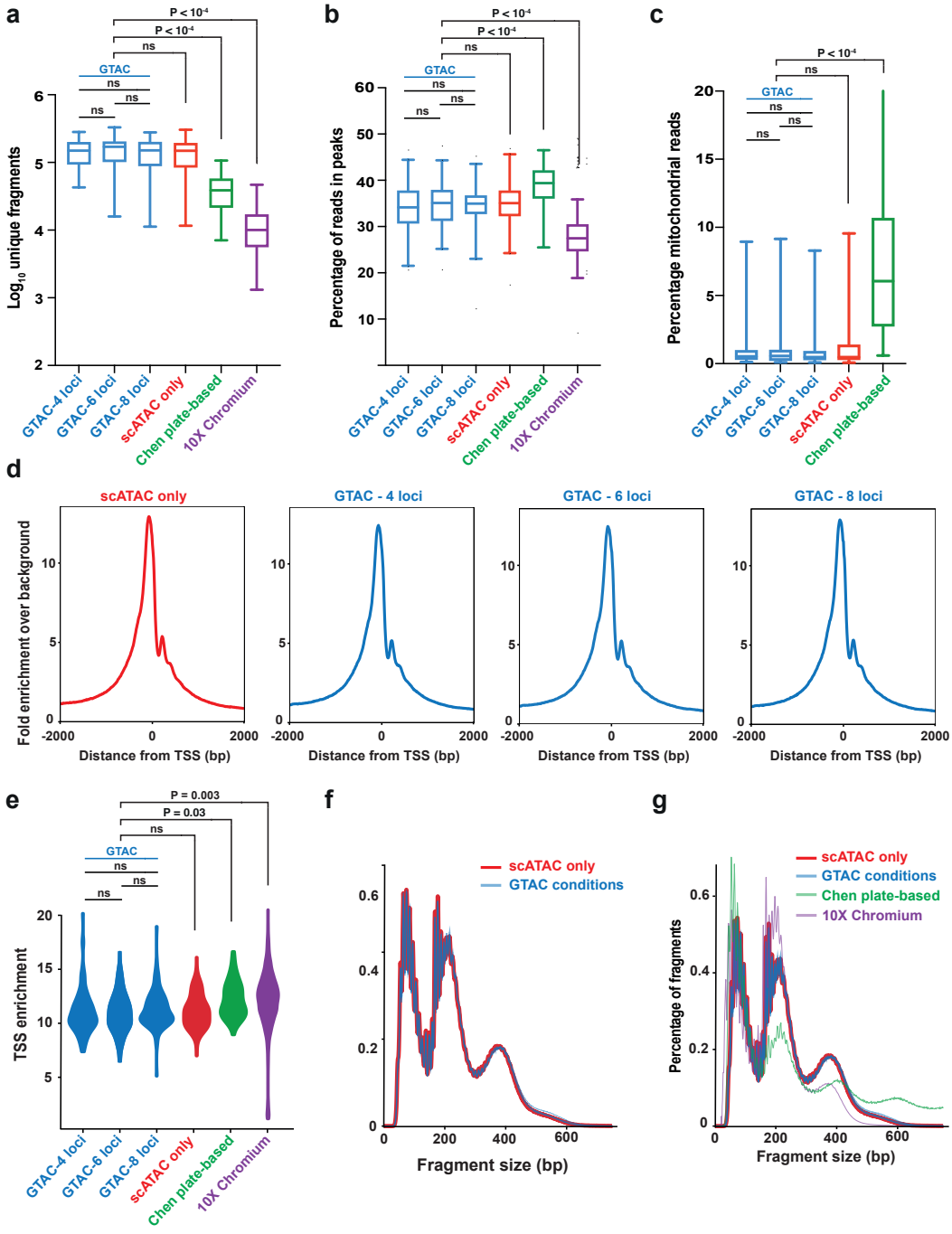


Figure 4.14: scATAC-seq library metrics in GTAC. In all panels, 4 conditions from Figure 4.13a are shown. In panels (a)-(c), (e), and (g), in-house scATAC-seq metrics were compared to published plate-based scATAC-seq K562 data (green, Chen plate-based); in panels (a), (b), (e), and (g), in-house metrics were additionally compared with published 10X Genomics scATAC-seq K562 data. In panels (a)-(c) and (e), ordinary one-way ANOVA with Dunnett's multiple comparisons was used. To compare GTAC metrics with standard scATAC-seq or published data, GTAC conditions were aggregated. **(a)** scATAC-seq library complexity across conditions. Median values are shown. Whiskers extend from the 2.5th to the 97.5th percentile. *Legend continues on next page.*

Figure 4.14 ***(previous page)*****: (b)** Percentage of scATAC-seq fragments mapping to peaks of open chromatin across conditions. The rest as in (a). **(c)** Percentage of mitochondrial reads across conditions. The rest as in (a). **(d)** Fold enrichment (y-axis) of scATAC-seq fragments mapped in proximity of transcription start sites (TSS) over the background across the genome, across in-house conditions. **(e)** As in (d), but quantifying per-cell TSS enrichment across conditions. **(f)** scATAC-seq fragment size distribution, aggregated across in-house conditions. **(g)** As in (f) but overlaid with public data.

The fraction of sequencing reads mapping to peaks of open chromatin (FRiP) measures signal-to-noise ratio (peaks of open chromatin are genomic regions with significantly higher signal over the background of the genome). FRiP was fully comparable between in-house GTAC and standard scATAC-seq conditions (Figure 4.14b). FRiP was subtly, yet significantly lower compared to previously published plate-based scATAC-seq, but higher compared to 10X Chromium. If the ATAC lysis steps are overly harsh, this can result in a significant overrepresentation of sequencing reads mapping to the mitochondrial genome. Consistent with high nuclear library complexity, the median percentage of mitochondrial reads was $< 1\%$ in all in-house GTAC and standard scATAC-seq conditions, significantly lower than in the prior plate-based method (Figure 4.14c). Moreover, in a successful ATAC experiment, an enrichment of fragments mapping to cis-regulatory elements over the background across the genome is expected. Failure to observe such enrichment may indicate that the native chromatin conformation was disrupted during lysis and tagmentation. To evaluate this, a measure of transcriptional start site (TSS) enrichment is frequently used. TSS enrichment was unaltered between in-house conditions, and substantially above the commonly adopted threshold of 4 [355] (Figures 4.14d-e). Of note, TSS enrichment was subtly lower compared to prior data (Figure 4.14e). I hypothesize this to be due to more efficient capture of di-nucleosomal fragments with my conditions (Figures 4.14f-g), whereas eukaryotic TSS-proximal promoter regions are predominantly nucleosome-free [392]. Importantly, fragment size distribution was unaltered between GTAC and standard scATAC-seq (Figure 4.14f). Open chromatin peak calls correlated well between in-house GTAC conditions and standard scATAC-seq ($r = 0.96$ in all comparisons; Figure 4.15a).

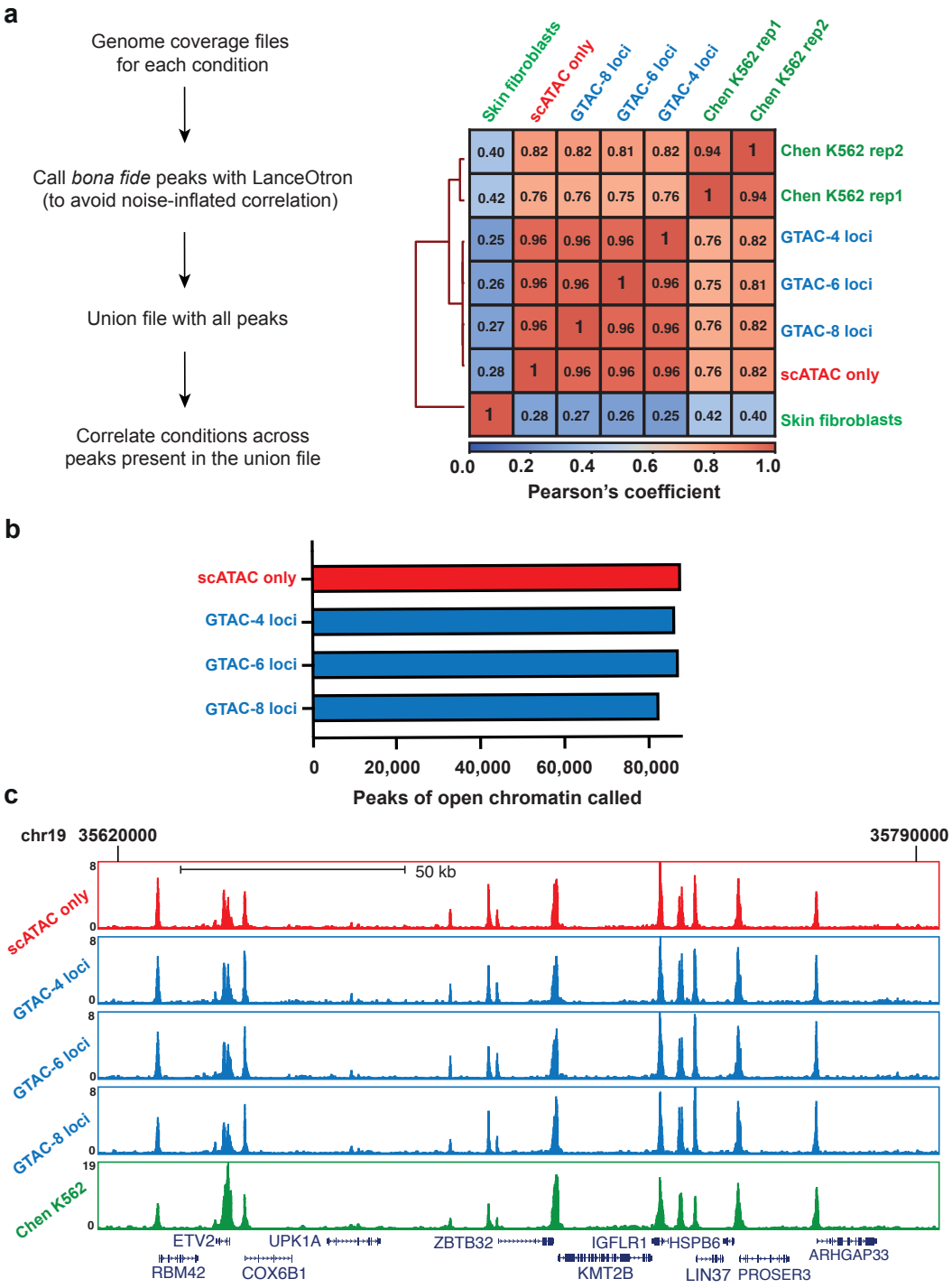


Figure 4.15: Correlation of scATAC-seq libraries. **(a)** Left, strategy for correlation of genomic peak calls across conditions. Right, matrix based on Pearson's coefficients, plotting correlation between in-house conditions and published plate-based scATAC-seq data from either K562 (two replicates) or human skin fibroblasts [349]. **(b)** Number of open chromatin peaks called for in-house conditions (80 K562 and 15 THP1 cells in each condition). **(c)** Aggregated scATAC-seq fragments (y-axis) from in-house conditions and published plate-based scATAC-seq K562 data.

In-house conditions also correlated well with two published plate-based scATAC-seq K562 datasets [349]. A stronger correlation between in-house conditions compared to correlation with public K562 data is likely due to batch effects between laboratories. Accordingly, the number of open chromatin peaks was comparable when peaks were called on the same number of cells in each in-house condition (Figure 4.15b). Visual inspection of genomic coverage across a pre-selected genomic locus suggested no obvious qualitative differences between conditions, as peaks were located within same genomic intervals (Figure 4.15c). Importantly, genomic coverage in proximity of the genotyped mutant/variant loci in K562 cells (Figure 4.16) showed no qualitative difference between GTAC and scATAC-seq. This (a) suggests there is no disruption to open chromatin fragment amplification in proximity of genotyped loci, and (b) confirms the genotyped genomic loci were not located in open chromatin, explaining failed capture in standard scATAC-seq (Figure 4.9b).

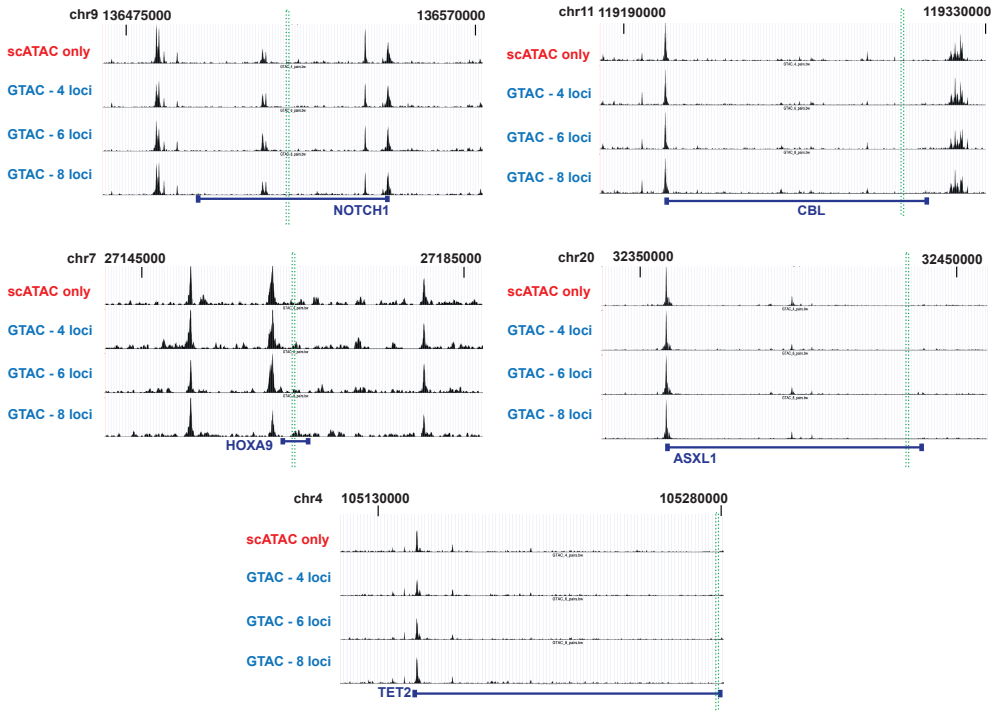


Figure 4.16: scATAC-seq coverage at genotyped loci. Aggregated scATAC-seq fragments (y-axis) from in-house conditions at 5 loci genotyped by GTAC in K562 cells (*NOTCH1*, *CBL*, *HOXA9*, *ASXL1*, *TET2*). Green rectangles indicate the rough position of the genotyping amplicon.

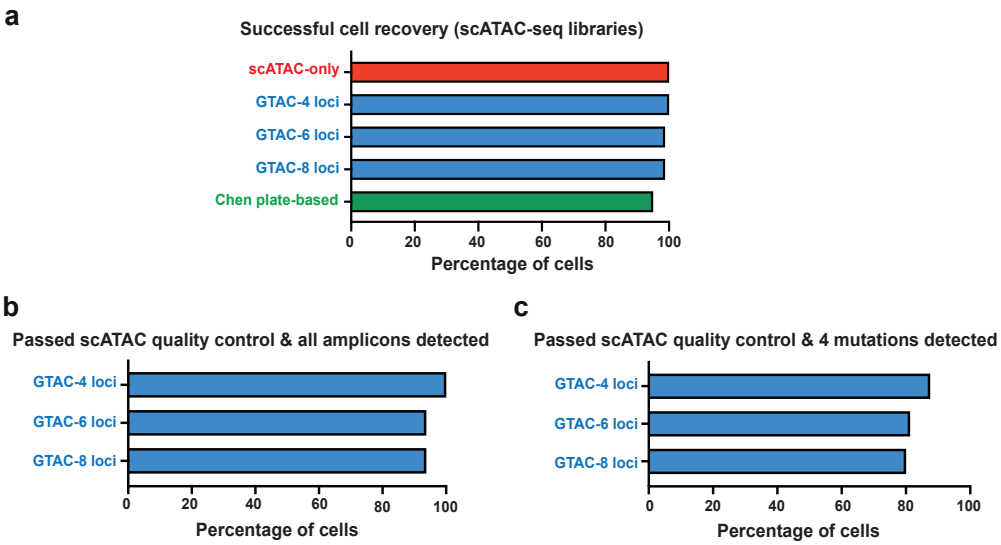


Figure 4.17: K562 cells passing quality filters in GTAC. (a) Cells passing scATAC quality filters (x-axis) across in-house conditions and published plate-based scATAC-seq K562 data. **(b)** Cells across GTAC conditions passing scATAC-seq quality filters and displaying detection of all genotyping loci. **(c)** Cells across GTAC conditions passing scATAC-seq quality filters, in which 4 heterozygous mutations were detected.

Importantly, > 90% of cells passed scATAC-seq quality control (QC; Materials and Methods) in all GTAC and standard scATAC-seq conditions (Figure 4.17a). Similarly, > 90% of cells in all GTAC conditions passed scATAC-seq QC and displayed successful capture of all genotyping amplicons (Figure 4.17b). Most importantly, > 80% of cells in all GTAC conditions passed scATAC-seq QC and displayed correct detection of all screened heterozygous mutations (Figure 4.17c). In summary, GTAC produces high quality scATAC-seq libraries, regardless of the number of genotyped loci, and most cells pass QC filters.

4.3.5 Integration of genotypes and chromatin accessibility profiles in cell lines

The main aim of GTAC is to compare chromatin accessibility between different genetic entities. Efficient integration of scATAC-seq and genotyping data was tested by a cell line mixing experiment with K562, JURKAT, and OCI-AML3 cells, in which 3 genomic loci were co-amplified in each cell line (1 mutant and 2 WT loci in each cell line). UMAP visualization efficiently distinguished the three cell lines based exclusively on chromatin accessibility (Figure 4.18a). Expectedly, three clusters were

detected, corresponding to three distinct cell lines. Importantly, mutations were called correctly in 97.2% of all cells and cell lines displayed the correct genotypes, with minimal false negative rates (up to 4%; Figure 4.18b). Hence, genotyping data is efficiently integrated with chromatin accessibility.

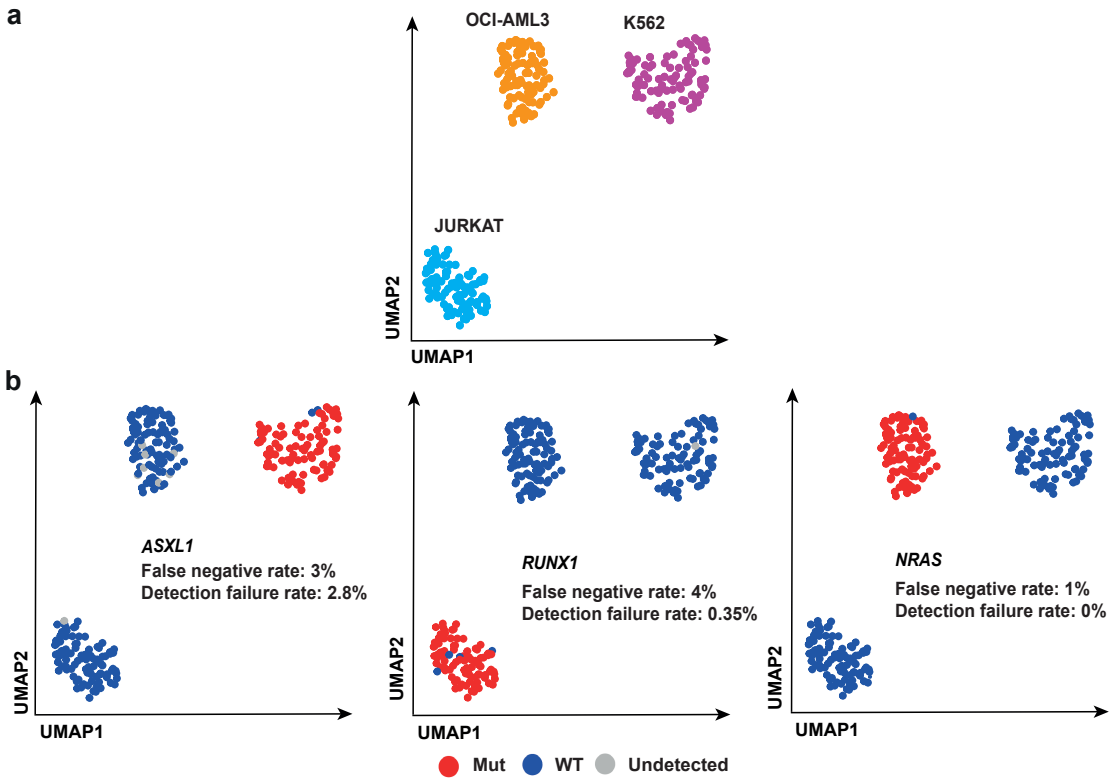


Figure 4.18: Cell line mixing experiment. K562, JURKAT and OCI-AML-3 cells (n = 95 each) were processed with GTAC in which *ASXL1*, *RUNX1*, and *NRAS* loci were genotyped in each cell line. **(a)** UMAP based on scATAC-seq data. **(b)** The UMAP from (a) overlaid with genotyping data. The mutation was either detected (red), not detected (blue), or single-cell genotyping failed (grey). The percentage of cells that are false negative (i.e. mutant ADO), or in which single-cell genotyping failed (Detection failure rate), is shown.

4.4 Brief summary of Chapter 4

The results in Chapter 4 validated the feasibility and performance of GTAC, a novel method for joint capture of accessible chromatin and mutational status at single-cell resolution. GTAC allows for high-confidence, accurate, and multiplexed genotyping of most single cells whilst providing robust and high-quality scATAC-seq libraries.

Chapter 5

GTAC links clonal evolution to epigenetic regulation in human AML

Contents

5.1 The main questions to address by using GTAC on human AML 127
5.1.1 GTAC is feasible in human BM 128
5.2 Execution of GTAC in human AML 130
5.2.1 Construction of scATAC-seq and single-cell genotyping libraries 130
5.2.2 Single-cell genotyping and scATAC-seq library metrics . 132
5.3 Integration of chromatin accessibility and genetic clonal identity in human AML 135
5.3.1 Inference of the most likely clonal hierarchy 135
5.3.2 The chromatin accessibility landscape in primary human AML 137
5.3.3 Integration of clonal identity with chromatin accessibility 144
5.3.4 Epigenomic correlates of *RUNX1*-R201G-associated leukaemic transformation 148
5.4 Consequences of subclonal *BCOR* mutations on leukaemic differentiation 155
5.4.1 Clonal basis for the LSC-like chromatin state within the leukaemic hierarchy 156
5.4.2 Differentiation delay of *BCOR*-mutant leukaemic cells . 159
5.5 Putative epigenetic mechanisms underlying selection of *BCOR*-mutant clones 164
5.5.1 Epigenetic consequences of *BCOR* mutations within the LMPP-like cluster 164
5.5.2 *BCOR* mutations de-regulate promoters of LSC-associated genes 170
5.5.3 Aberrant TF activity in *BCOR*-mutant LMPP-like cells 173
5.5.4 Epigenetic heterogeneity between genetically similar clones 177
5.6 Brief summary of Chapter 5 179

I am grateful to Angus Groom, for precious help with GTAC primer testing and processing of single-cell genotyping plates. Most of the work in this Chapter is adapted from its original publication in Cell Stem Cell [332] (Appendix A.3)

5.1 The main questions to address by using GTAC on human AML

Having benchmarked the technical aspects of GTAC, I next employed this method to decipher how cooperating somatic mutations corrupt gene regulation in human AML. The objective was two-fold: (a) to demonstrate that GTAC efficiently links genetic clonal hierarchy and chromatin accessibility in primary human tissue; and (b) to provide biologically relevant insights into aberrant gene regulation downstream of specific somatic events. AML was an adequate disease to interrogate by GTAC, for several reasons:

1. Patient outcome is still dismal and relapse is frequent, indicating that our knowledge of how somatic lesions support clonal advantage is incomplete.

2. Genes encoding TFs and epigenetic modifiers are frequently mutated in AML, but the downstream effects of these mutations are poorly understood, potentially due to the pleiotropic roles these proteins exert on chromatin. In this regard, a method for global analysis of chromatin landscapes downstream of these mutations may lead to mechanistic insights about their consequences.

3. Treatment failure in AML is linked to the persistence of LSCs. Cooperating somatic lesions reprogram gene regulation in normal progenitors/stem cells, converting them into LSCs. Understanding this process is critical for the development of LSC-targeted therapies.

4. AML is one of the least mutated cancers [173]. This allows to benchmark GTAC by screening a tractable number of somatic lesions, without confounding effects of a complex somatic background.

Therefore, I applied GTAC to a diagnostic AML patient sample harbouring multiple mutations in genes encoding epigenetic regulators and transcription factors, namely *TET2*, *ASXL1*, *RUNX1*, and *BCOR*. Despite their recurrence in myeloid

malignancy, it is not completely understood how these mutations promote malignant phenotypes. Whilst mutations in *TET2* and *ASXL1* are early pre-leukaemic events, *RUNX1* and *BCOR* mutations tend to occur as later events, suggesting transforming potential of these mutations [172, 281, 393, 394]. Moreover, mutations in *ASXL1*, *RUNX1*, and *BCOR* confer poor prognosis [172]. Interestingly, *RUNX1* and *BCOR* mutations frequently co-occur in AML [297–299], suggesting mutation cooperativity, but the molecular reason for this is unclear. I asked the following questions:

1. Which mutation was the major transforming event? How did driver mutations modify chromatin landscapes to convert pre-leukaemic into leukaemic cells?
2. How did mutations alter chromatin to support LSC states?
3. What is the mechanistic basis for *RUNX1* and *BCOR* mutation cooperativity?

Chapter 5 first demonstrates that using GTAC on primary human BM is feasible, with minor modifications to the protocol (section 5.1.1). Next, GTAC is applied to the primary AML sample, revealing how the gradual acquisition of somatic mutations reshapes epigenetic landscapes.

5.1.1 GTAC is feasible in human BM

I first optimized GTAC for application on human BM material. For this, I used a non-leukaemic BM sample from the MARCH study [80]. $CD34^+$ cells were pre-sorted (FACS panel as in Figure 3.2a but DAPI was used instead of 7-AAD), lysis and Tn5 tagmentation were performed, and single nuclei were FACS-sorted (strategy as in Figure 4.4b). qPCR indicated that 16 PCR cycles were optimal for human BM $CD34^+$ cells (Figure 5.1a). Moreover, a PCR elongation time of 30 s was optimal to generate scATAC-seq libraries displaying the nucleosomal pattern (Figure 5.1b). Using these conditions, GTAC was tested on human BM $CD34^+$ cells, adding two genotyping primer pairs against *DNMT3A* and *TET2* loci. The scATAC-seq library and both genotyping amplicons were successfully generated (Figure 5.1c), demonstrating the feasibility of GTAC in primary human BM. Next,

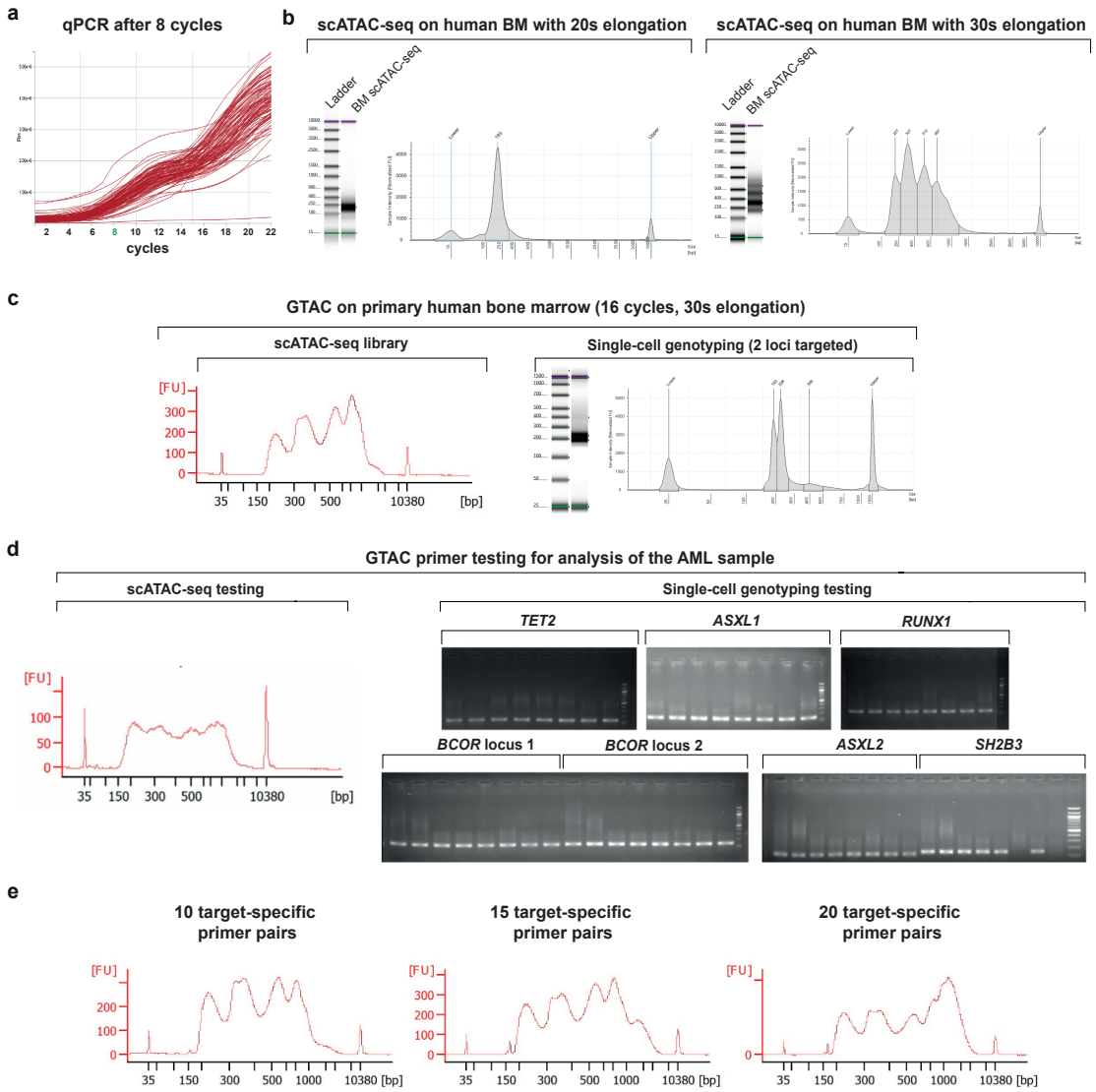


Figure 5.1: GTAC testing in primary human BM. **(a)** qPCR plot as in Figure 4.4c but from single sorted CD34$^+$ primary human BM cells. **(b)** Pooled scATAC-seq libraries (standard scATAC on 382 single cells) from human BM CD34$^+$ cells, when PCR elongation time was 20s (left) or 30s (right). **(c)** GTAC on primary human BM CD34$^+$ cells. Left, scATAC-seq library from pooled 382 single cells. Right, two genotyping amplicons of the expected size from the same cell pool. **(d)** Single-cell GTAC testing of 7 primer pairs to be used for analysis of the selected AML sample, performed using human BM CD34$^+$ cells. Left, scATAC trace (24 pooled single cells). Right, single-cell genotyping amplicons for the 7 loci. **(e)** scATAC-seq traces from a GTAC experiment in which 10, 15, or 20 loci were co-amplified (pools of 96 primary human BM CD34$^+$ cells). Panel (e) is adapted from [333].

using the framework described in Figure 4.8 on human CD34$^+$ BM cells, I tested primers against 7 loci to be genotyped in the primary human AML sample. This identified 7 primer pairs which generated genotyping amplicons of correct size in single cells whilst preserving scATAC-seq library shape (Figure 5.1d and Table A.13). Finally, although I rigorously benchmarked GTAC using a maximum of 8 genotyping primer pairs (Chapter 4), for some applications, it may be necessary to genotype more genomic loci. Encouragingly, scATAC-seq library shape derived from pooled single human BM CD34$^+$ cells was qualitatively unaltered when 10, 15, or even 20 genotyping primer pairs were added to the PCR (Figure 5.1e). This suggests no disruption to human scATAC-seq libraries when higher numbers of loci are genotyped. Overall, GTAC is applicable to human BM.

5.2 Execution of GTAC in human AML

5.2.1 Construction of scATAC-seq and single-cell genotyping libraries

Using the experimental conditions established above (Figure 5.1), I applied GTAC to 3,520 single cells from a human AML diagnostic patient sample harbouring *TET2*, *ASXL1*, *RUNX1*, and two *BCOR* mutations (Figure 5.2a and Table 5.1). In addition, the patient harboured heterozygous germline SNPs in *ASXL2* and *SH2B3* (Table 5.1). Somatic mutations and germline SNPs were genotyped in the GTAC experiment, by using 7 genotyping primer pairs (Figure 5.1d and Table A.13). Whilst somatic mutations were genotyped to establish the clonal hierarchy and assign each cell to a clone, germline SNPs were used for ADO rate estimation.

Gene	Exon	DNA_change	Protein_change	Chromosome	Status
BCOR	exon4	c.2607_2608insCC	p.T870fs	chrX	Somatic
BCOR	exon4	c.1495_1496insAGCG	p.S499_E500delinsX	chrX	Somatic
TET2	exon6	c.3626T>C	p.L1209P	chr4	Somatic
ASXL1	exon12	c.2061T>A	p.C687X	chr20	Somatic
RUNX1	exon6	c.601G>C	p.R201G	chr21	Somatic
ASXL2	exon12	c.3600G>T	p.Q1200H	chr2	Germline
SH2B3	exon6	c.1183G>A	p.E395K	chr12	Germline

Table 5.1: Somatic mutations and germline heterozygous SNPs genotyped by GTAC in the primary human AML sample. The targeted NGS panel was performed by Ms. Marlen Metzner.

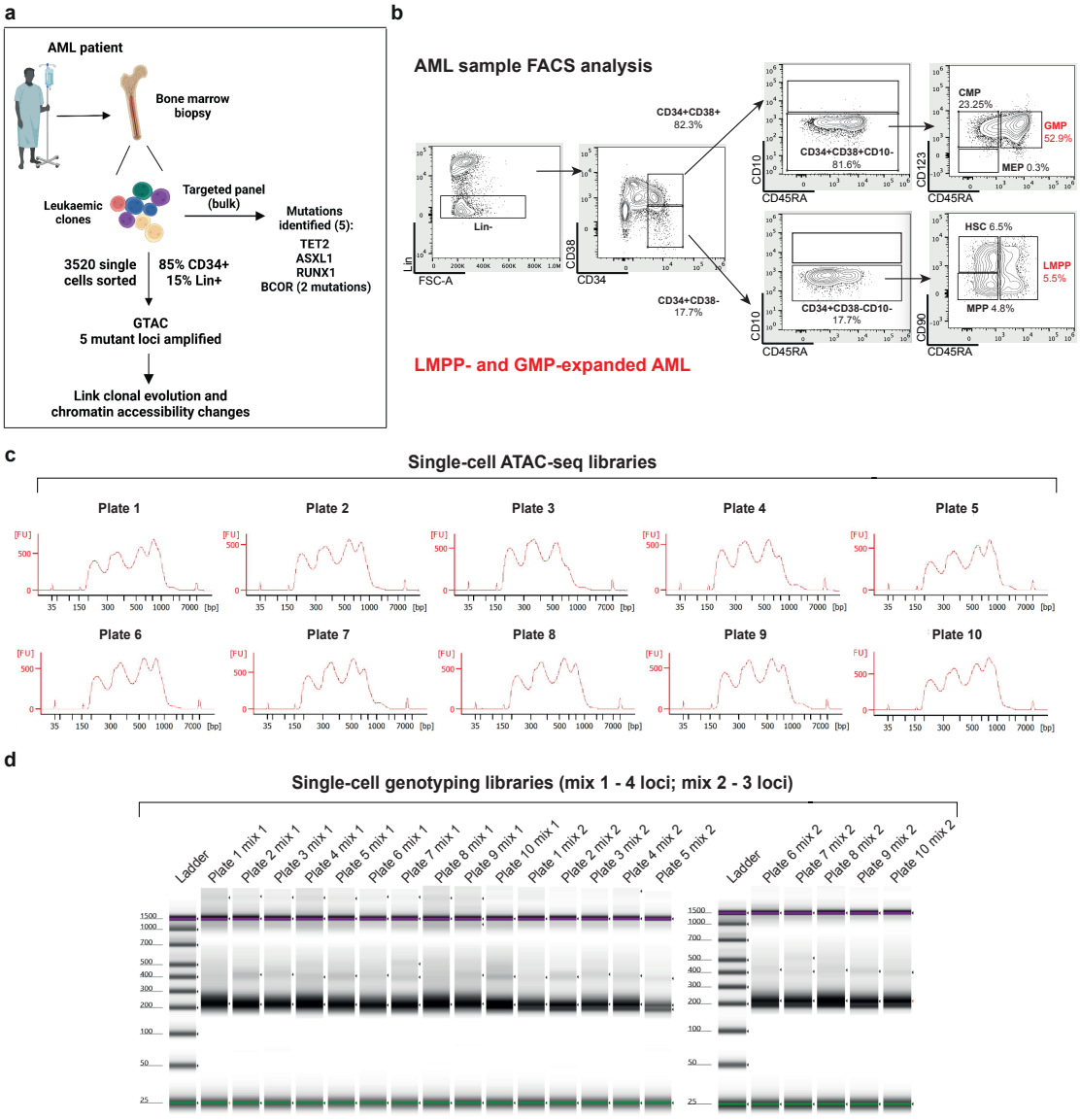


Figure 5.2: GTAC on primary human AML. (a) GTAC was performed on 3,520 single cells (≈85% $CD34^+$ and ≈15% $Lineage^+$) from a male AML patient in which 5 driver mutations were identified by bulk NGS and were thus genotyped in GTAC. **(b)** FACS analysis of the patient sample from (a). The contribution (percentage) of each HSPC sub-population to all Lin^-CD34^+ HSPCs is indicated. **(c)** scATAC-seq libraries from each processed 384-well plate. **(d)** Genotyping libraries from each processed 384-well plate (2 genotyping mixes per plate, containing 4 and 3 amplicons).

As control for single-cell genotyping (both for mutations and for SNPs), 300 single cells from a healthy human BM sample, WT at all genotyped loci, were also subject to GTAC. FACS analysis of the AML sample revealed aberrantly expanded GMP and LMPP compartments (Figure 5.2b). Lysis and tagmentation were performed on a pre-sorted cell mix, containing ≈85% of $CD34^+$ and ≈15% of Lin^+ cells. Nuclei were sorted into 10 384-well plates and both scATAC-seq (Figure 5.2c) and single-cell genotyping (Figure 5.2d) libraries were successfully generated from all plates, with no apparent technical differences between plates.

5.2.2 Single-cell genotyping and scATAC-seq library metrics

I first estimated ADO rates in the AML sample. Whilst it was not possible to directly measure ADO for the 5 somatic mutations, an ADO estimate was obtained via single-cell genotyping of the 2 heterozygous SNPs (Figure 5.3a). scVAF distributions at both loci were consistent with a heterozygous state (median 0.56 for *SH2B3* and 0.52 for *ASXL2*, Figure 5.3b). There was more scVAF spread at the *SH2B3* locus, indicating a degree of variability in noise levels across loci. Next, thresholds for heterozygous and homozygous variant calling (as in Figure 4.11a) were set using WT control BM cells (Figure 5.3c), allowing to estimate ADO rates for the variant and reference alleles. This demonstrated low variant (2.2% and 3.1%) and reference (2.0% and 3.4%) ADO percentages at *SH2B3* and *ASXL2* loci, respectively (Figure 5.3d). Consequently, bi-allelic detection occurred in 95.8% of cells for the *SH2B3* SNP and in 93.5% of cells for the *ASXL2* SNP (Figure 5.3d). Importantly, rates of bi-allelic detection at both loci were comparable between plates (Figure 5.3e), indicating no variations in single-cell genotyping accuracy between plates.

Next, I benchmarked scATAC-seq library quality. As comparison, I used scATAC-seq metrics from 5 previously published [88] mixed-phenotype acute leukaemia (MPAL) human samples processed with 10X Chromium (the AML sample processed in-house is likely to derive from an MPAL patient, detailed below). Library complexity was high, with a median of 70,475 unique nuclear fragments per cell, compared to 12,192 in 10X Chromium (Figure 5.4a).

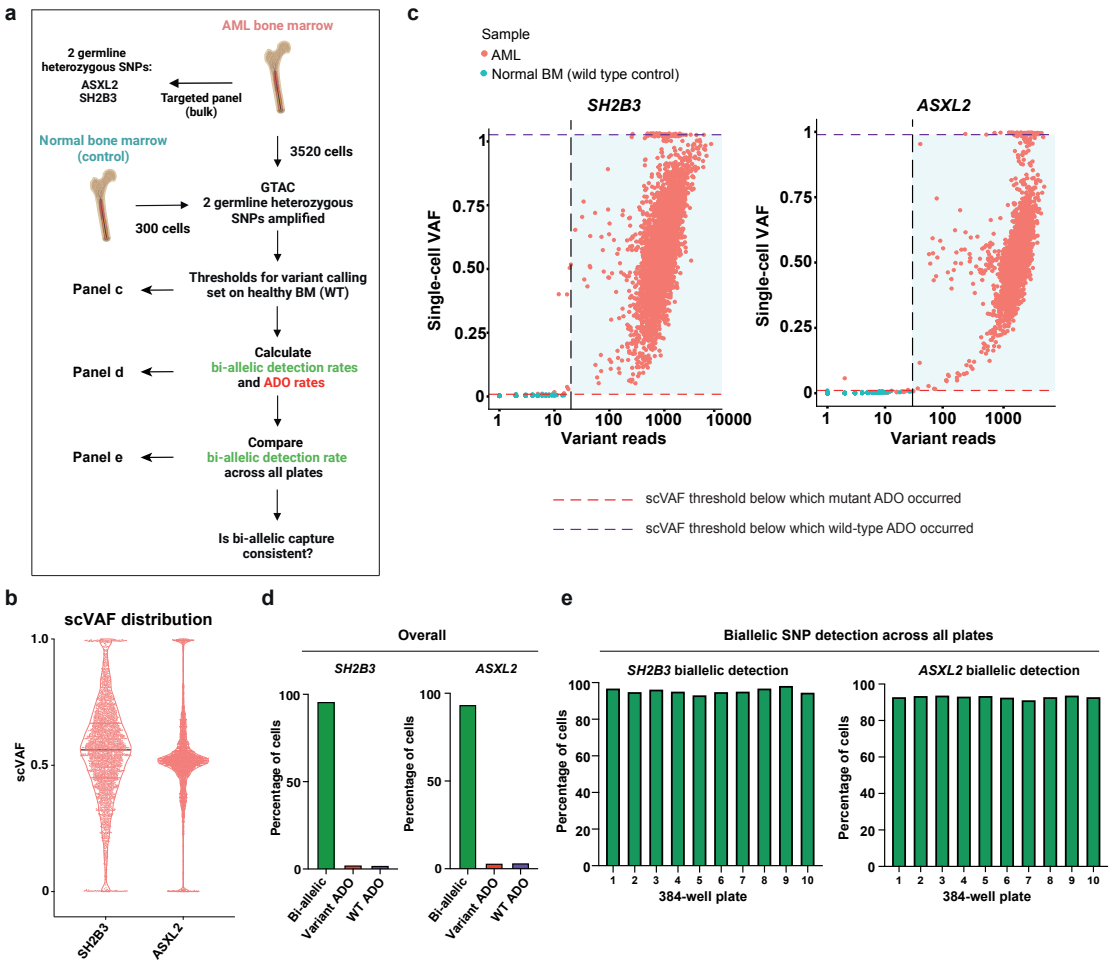


Figure 5.3: ADO estimation in the AML sample. **(a)** Two germline heterozygous SNPs in *ASXL2* and *SH2B3* genes were detected by bulk NGS and amplified in GTAC, in the AML sample and in 300 CD34$^+$ cells from a WT control human BM. **(b)** scVAF distribution for the two SNPs in the AML sample. Median values shown. **(c)** scVAF (y-axis) versus the number of variant reads (x-axis) for the heterozygous germline SNPs (indicated above). AML patient cells are in red. Healthy CD34$^+$ BM cells (cyan) were used as WT control. Dashed lines indicate genotyping thresholds (sections 2.11.1 and 2.11.2). ADO occurred in those AML cells outside the azure rectangle. **(d)** Percentage of AML cells showing bi-allelic detection (green; i.e. there was no ADO), or in which variant ADO (red) or WT ADO (purple) occurred, at the 2 SNPs. Bi-allelic detection/variant ADO/WT ADO percentages: *SH2B3*: 95.8/2.2/2.0; *ASXL2*: 93.5/3.1/3.4. **(e)** Bi-allelic detection percentage for the two SNPs across plates.

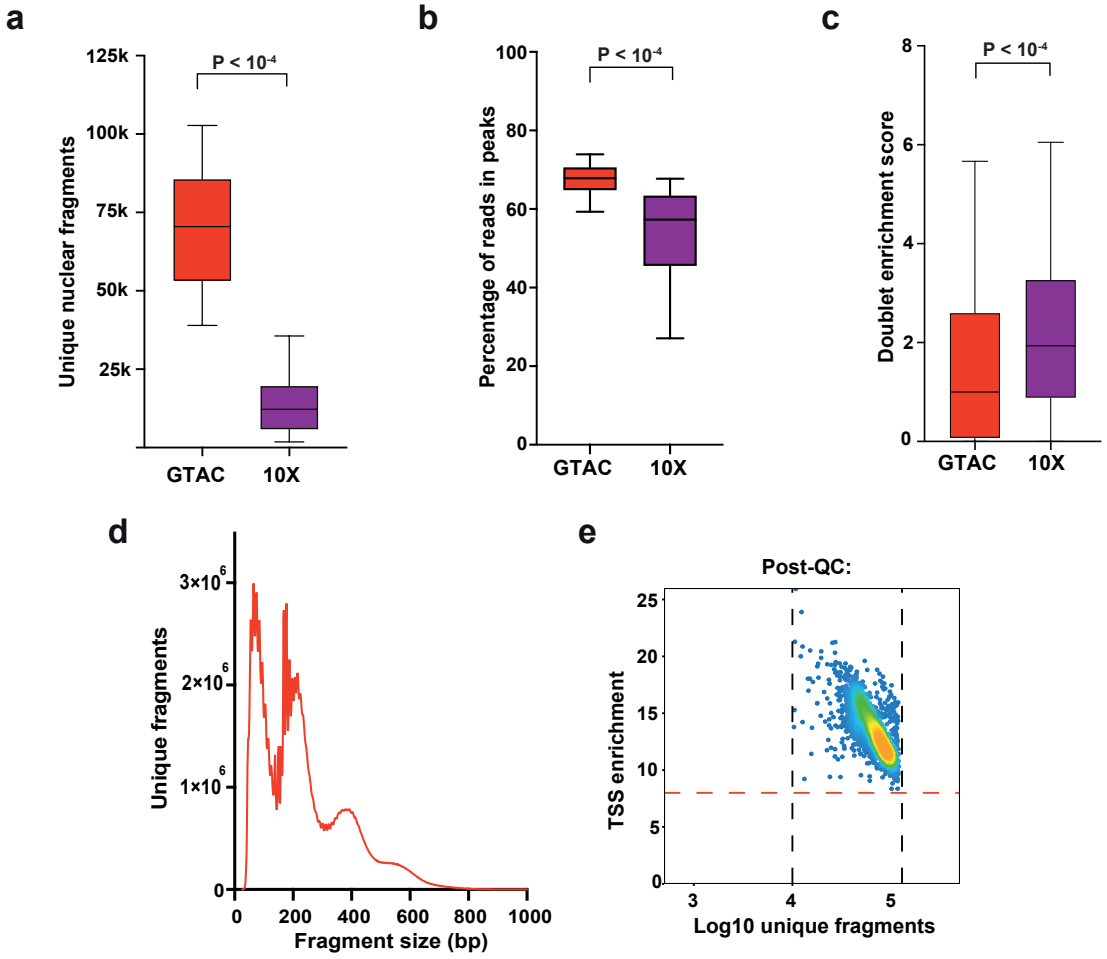


Figure 5.4: scATAC-seq metrics in the AML sample. In panels (a)-(c), scATAC-seq metrics from the AML sample processed with GTAC (red, n = 3,285 cells) were compared to public data derived from 5 MPAL samples processed with 10X Chromium (purple, n = 20,059 cells). In both datasets, cells with > 1,000 and < 120,000 unique nuclear fragments were used. **(a)** scATAC-seq library complexity. Median number of unique fragments: 70,475 for GTAC, 12,192 for 10X. A two-tailed t-test was used. **(b)** As in (a) but showing the percentage of unique scATAC-seq fragments mapping to peaks of open chromatin. Median values: 68% for GTAC, 57% for 10X. **(c)** As in (a) but showing doublet enrichment scores. Median values: 1.00 for GTAC, 1.93 for 10X. **(d)** Aggregated scATAC-seq fragment size distribution for the AML sample processed with GTAC. **(e)** Log_{10} of the number of unique nuclear fragments (x-axis) plotted against the TSS enrichment score (y-axis) for AML cells (dots) processed with GTAC. Cells passing QC (section 2.11.4) were included.

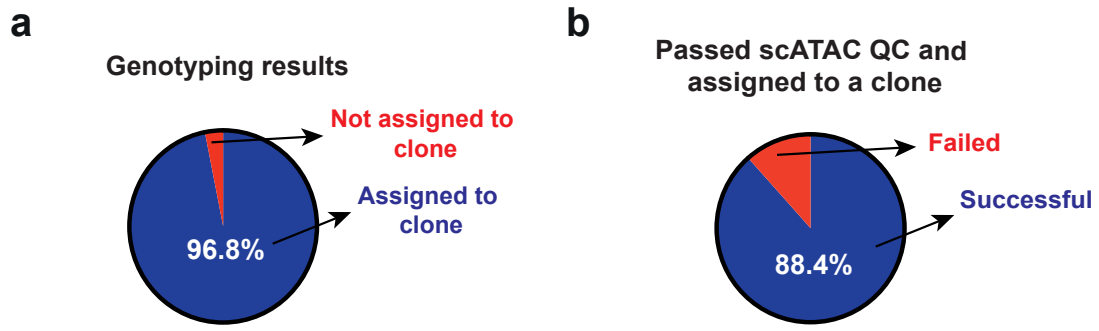


Figure 5.5: AML cells passing GTAC quality filters. (a) Percentage of AML cells assigned to a genetic clone (3,407/3,520 cells). **(b)** Percentage of AML cells which passed scATAC-seq quality filters and which were assigned to a clone (3,113/3,520 cells).

FRiP was 68%, indicating elevated signal-to-noise ratio, significantly higher compared to 10X Chromium (Figure 5.4b). The inferred doublet rate was significantly lower in GTAC (Figure 5.4c), likely due to doublet exclusion via FACS. Fragment size distribution was as expected from a high-quality scATAC-seq library (Figure 5.4d). Consistently, cells passing scATAC-seq QC had high TSS enrichment scores (Figure 5.4e). Altogether, this suggests generation of interpretable chromatin accessibility data.

Despite lower throughput compared to droplet-based technologies, most cells processed with GTAC successfully passed quality filters. Indeed, 3,409 single cells (96.8% of all sorted cells) were assigned to genetic clones (Figure 5.5a). When these cells were filtered for those successfully passing conservative scATAC-seq QC filters (section 2.11.4), 3,113 single cells (88.4% of all sorted cells) passed QC (Figure 5.5b) and were used for downstream analyses.

5.3 Integration of chromatin accessibility and genetic clonal identity in human AML

5.3.1 Inference of the most likely clonal hierarchy

To establish the most likely course of somatic events and reconstruct the genetic clonal hierarchy, I called single-cell genotypes at the 5 mutant loci. As for SNPs, scVAFs of WT control BM cells were used to set thresholds for mutation calling (Figure 5.6a). The resulting scVAF distributions expectedly revealed a heterozygous

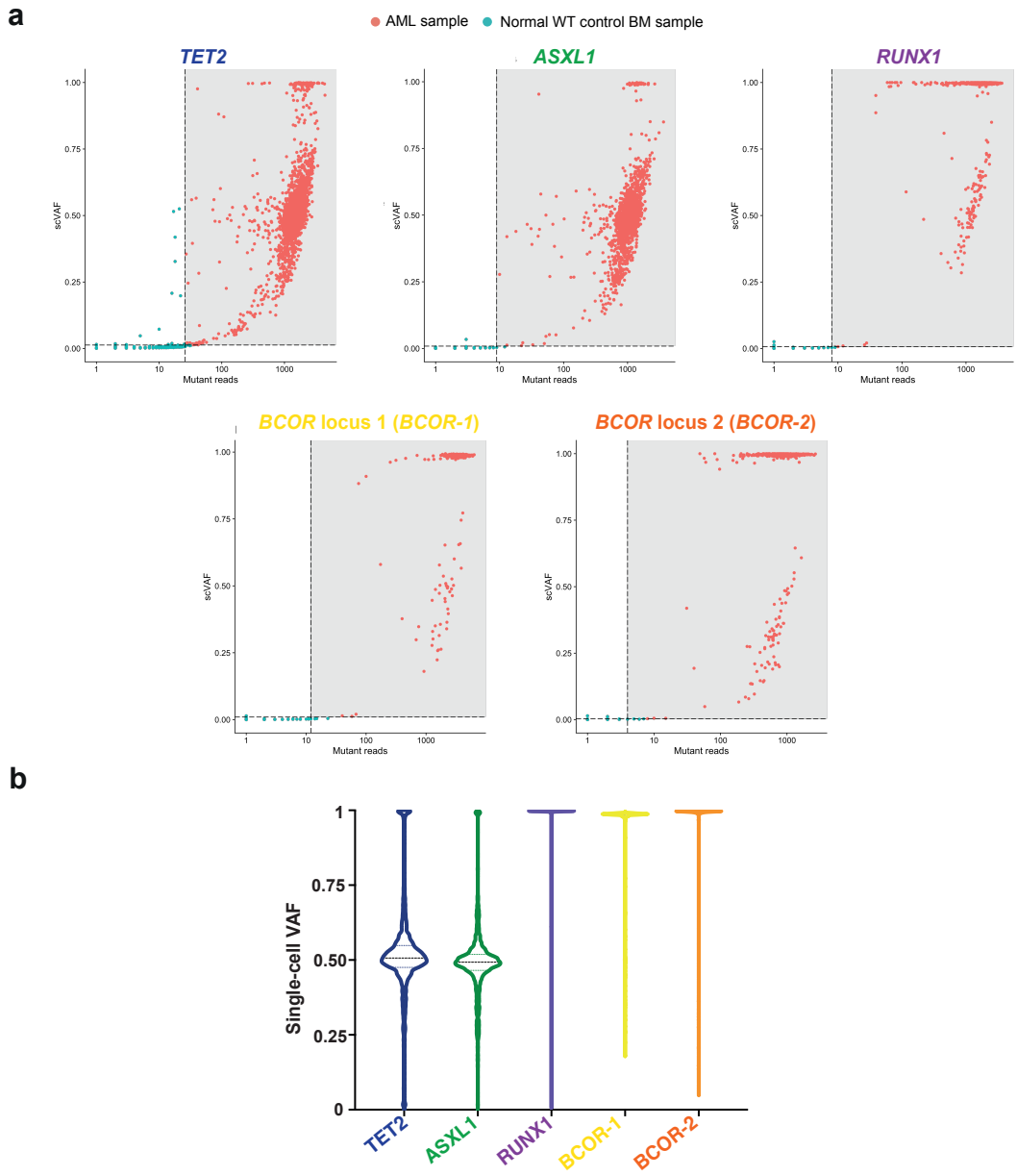


Figure 5.6: Single-cell genotyping of the AML sample. (a) Same as in Figure 5.3c but for the 5 genotyped somatic mutations. Cells within the grey rectangle are called mutant. **(b)** scVAF distributions in cells called mutant at the respective genomic loci. '*BCOR*-1' and '*BCOR*-2' indicate two different *BCOR* somatic mutations. Median scVAFs: *TET2*, 0.507; *ASXL1*, 0.493; *RUNX1*, 0.997; *BCOR*-1, 0.987; *BCOR*-2, 0.996.

state for *TET2* and *ASXL1* mutations (Figure 5.6b). A small proportion of apparently homozygous mutant cells was most likely due to WT ADO (3.6% for *TET2* and 1.5% for *ASXL1*). In contrast, the majority of *RUNX1*-mutant cells harboured a homozygous *RUNX1* mutation (Figure 5.6b). In this case, I inferred most of these cells were *bona fide* homozygous mutant, as their proportion greatly

exceeded the WT ADO rates predicted for any locus (Figures 4.11b and 5.3). A minority of *RUNX1*-mutant cells displayed a heterozygous genotype: these cells either belonged to an ancestral heterozygous clone or were doublets (one mutant and one WT cell). The *RUNX1*-R201G missense mutation, harboured by this patient, is predicted to abolish RUNX1 DNA binding properties [395–397]. Hence, *RUNX1*-R201G homozygous mutant cells in this sample likely harbour a complete *RUNX1* LoF. The two X-linked *BCOR* mutations were hemizygous in most cells (Figure 5.6b), expected given the patient was male (the 'heterozygous' cells were almost certainly doublets). Given that both mutations were predicted LoF, *BCOR*-mutant cells in this sample harbour a complete *BCOR* LoF.

Next, for each cell, I integrated the mutational status across the 5 genotyped loci (Figure 5.7a), allowing to reconstruct the most likely course of mutation acquisition (Figure 5.7b). This revealed that *TET2* (T), *ASXL1* (A), and *RUNX1* (R) mutations occurred serially (in the T, TA, and TAR clones, respectively). The two *BCOR* mutations (B1 and B2) were then acquired independently by the TAR clone, generating two branched terminal clones (TARB1 and TARB2; Figure 5.7b).

Interestingly, integration of scATAC-seq data to clonal information revealed a substantial difference in nuclear library complexity between clones, suggesting substantial differences in the degree of open chromatin between clones (Figure 5.7c). I used clonal information to generate cell pseudobulks and call peaks of open chromatin. Consistent with higher library complexity, the T clone displayed the highest number of peaks (Figure 5.7d).

Overall, GTAC allowed to reconstruct the mutational clonal hierarchy, thus providing the genetic clonal identity for each single cell.

5.3.2 The chromatin accessibility landscape in primary human AML

To generate the epigenomic cellular landscape based on chromatin accessibility, I used the scATAC-seq data to perform dimensionality reduction, UMAP visualization, and clustering (Figure 5.8). The resulting cellular landscape was based exclusively on chromatin accessibility and was agnostic to genetic clonal identity. Chromatin

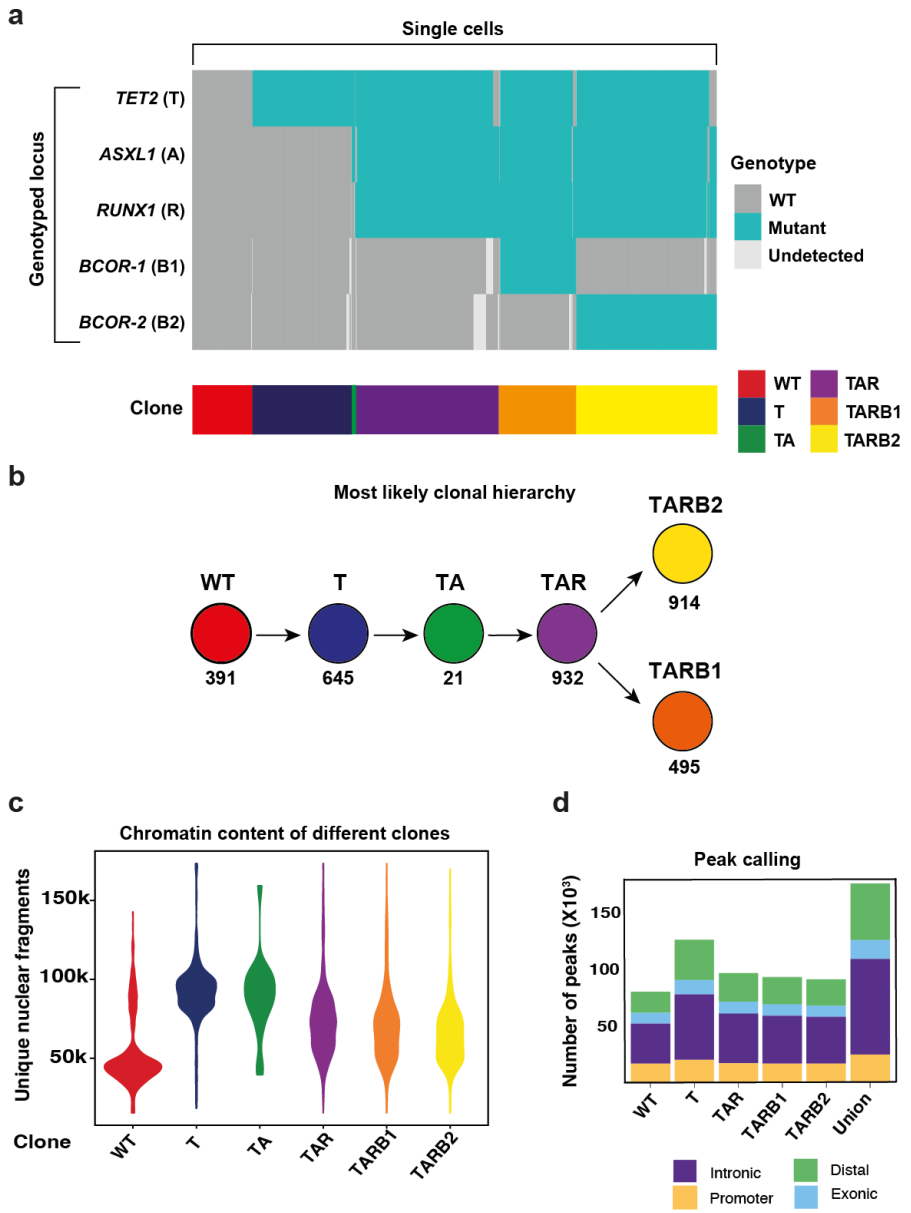


Figure 5.7: Clonal hierarchy reconstruction. **(a)** Above, single-cell genotyping results within all AML cells assigned to a clone (n = 3,407). Genotyped loci displayed as rows, single cells as columns. For each locus, cells are either WT (gray), mutant (cyan), or genotyping failed (Undetected, white). Genotyped mutations: *TET2*-L1209P (T), *ASXL1*-C687X (A), *RUNX1*-R201G (R), *BCOR*-T860fs (B1), and *BCOR*-S499_E500delinsX (B2). Below, single-cell clonal assignment. Clone abbreviations: WT, wild-type (red). T, *TET2*-mutant (dark blue). TA, *TET2*/*ASXL1*-mutant (green). TAR, *TET2*/*ASXL1*/*RUNX1*-mutant (purple). TARB1, *TET2*/*ASXL1*/*RUNX1*/*BCOR*-T860fs-mutant (orange). TARB2, *TET2*/*ASXL1*/*RUNX1*/*BCOR*-S499_E500delinsX-mutant (yellow). **(b)** Most probable clonal structure. **(c)** Clone-specific number of unique nuclear scATAC-seq fragments. **(d)** Clone-specific number of open chromatin peaks. Promoter peaks: 2 kb upstream and 100 bp downstream of annotated promoters.

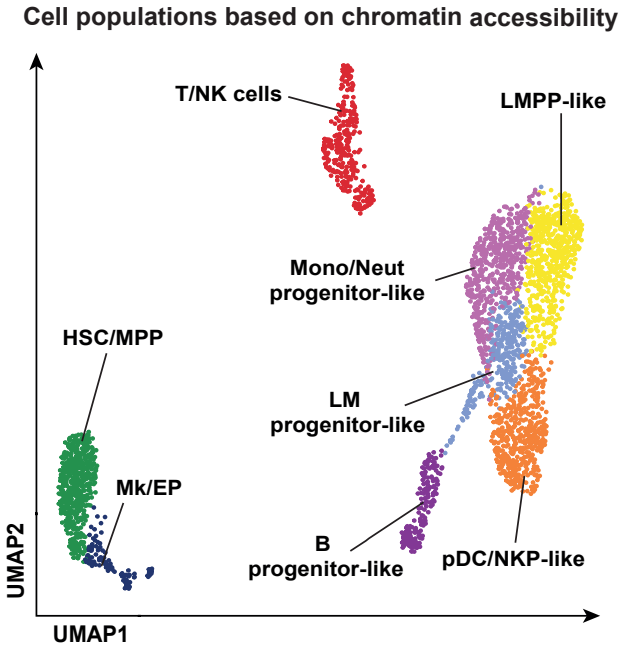


Figure 5.8: Chromatin accessibility landscape of the AML sample. UMAP based on scATAC-seq, depicting the chromatin accessibility landscape of 3,113 cells from the AML sample. HSC/MPP: hematopoietic stem cells/multipotent progenitors; Mk/EP: megakaryocytic/erythroid progenitors; LMPP-like: lymphoid-primed multipotent progenitor-like; Mono/Neut progenitor-like: monocyte/neutrophil progenitor-like; LM progenitor-like: lympho-myeloid progenitor-like; pDC/NKP-like: plasmacytoid dendritic cell/NK progenitor-like.

accessibility within each cell cluster informed cluster labelling. Briefly, clusters were assigned using gene activity (i.e. accessibility) of key marker genes (Figure 5.9a), enrichment of published haematopoietic signatures [81, 89, 90, 358–360] (Figure 5.9b), accessibility of TF binding motifs (Figure 5.9c), enrichment of open chromatin peaks in published bulk ATAC-seq signatures (Figure 5.9d), and mapping onto a healthy haematopoiesis scATAC-seq atlas [88] (Figures 5.10a-b). No plate-related batch effects were detected (Figure 5.9e), indicating clusters represented epigenetically distinct cellular entities rather than technical artefacts.

Within the chromatin landscape, 3 separate cellular entities were identified (Figure 5.8). One consisted of terminally mature T-cells and NK-cells (cluster: T/NK cells). These cells showed high accessibility of *CD3*, *CD8*, *GZMB*, and *TBX21* genes, enrichment of open chromatin for T/NK-cell transcriptomic/ATAC signatures, and high LEF1 binding motif accessibility (Figures 5.9a-d and 5.10c). The second

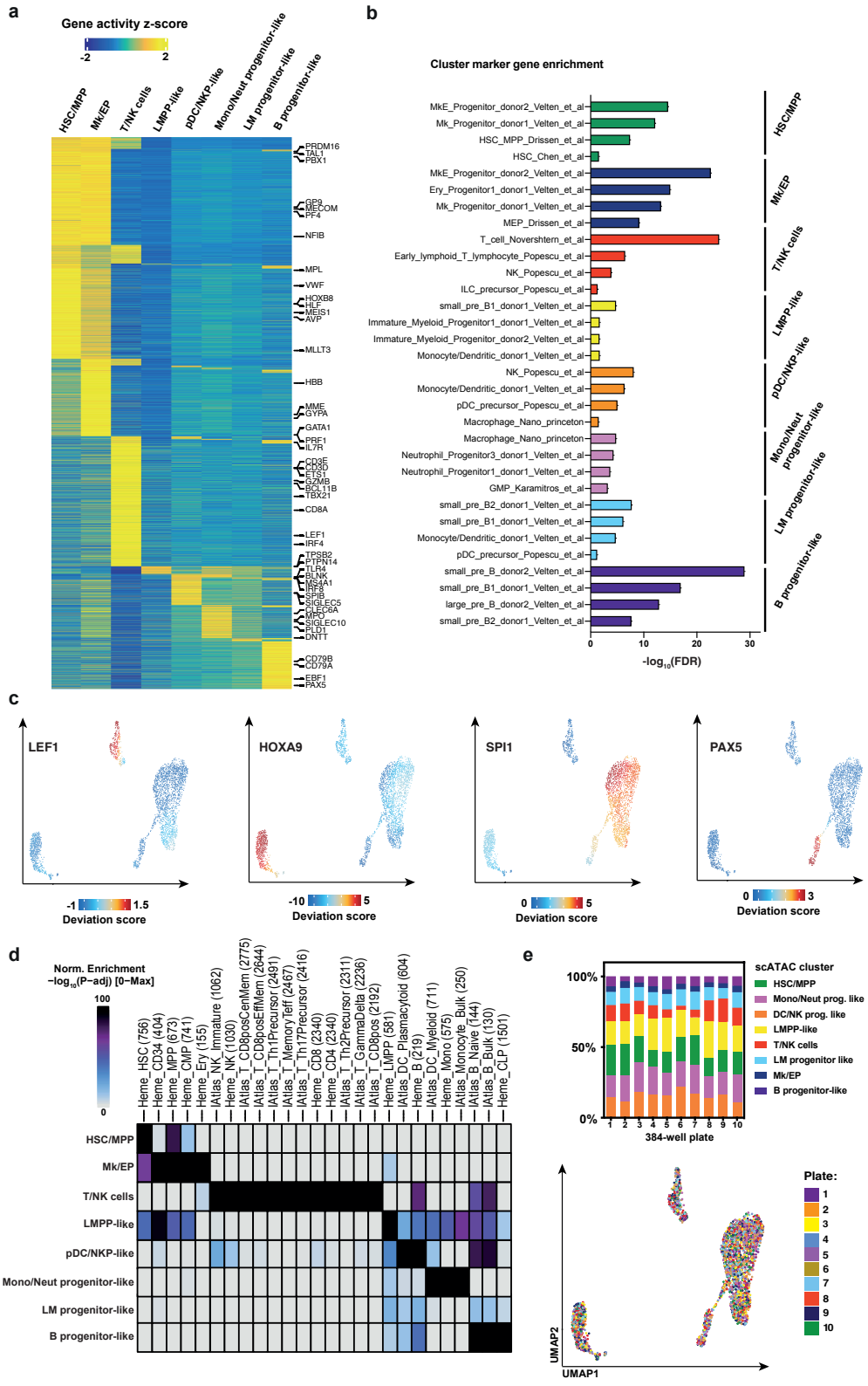


Figure 5.9: *(Legend on next page)*

Figure 5.9 ***(previous page)*****: scATAC-seq cluster annotation. (a)** Gene activity (i.e. accessibility) scores for cluster marker genes (rows), obtained when cells from one cluster were compared to all other clusters. Genes with FDR < 0.05 and $\log_2$FC > 0.5 are plotted. **(b)** Enrichment of cluster-specific accessible chromatin (marker genes with FDR < 0.05 and $\log_2$FC > 0.25) in haematopoietic transcriptional signatures. **(c)** The UMAP from Figure 5.8 labelled with chromVAR-computed accessibility of motifs bound by LEF1, HOXA9, SPI1, and PAX5. **(d)** Enrichment of cluster-specific marker peaks (FDR < 0.05 and $\log_2$FC > 0.25) in public bulk ATAC-seq signatures (provided by ArchR). Colours indicate normalized peak set enrichment scores (-$\log_{10}$(adjusted P-value)) on a scale of 0-100. The maximum -$\log_{10}$(adjusted P-value) for each set is indicated in brackets. **(e)** Evaluation of plate-related batch effects. Above: contribution of each scATAC-seq cluster (colours) to each 384-well plate (x-axis). Below: UMAP from Figure 5.8 labelled by 384-well plate.

population matched HSCs and early MPP populations (cluster: HSC/MPP), which formed a continuum with a smaller number of more committed progenitors, mainly from the megakaryocytic and erythroid lineages (cluster: Mk/EP). Accordingly, HSC/MPP cells had high accessibility of *HLF*, *AVP*, *MECOM*, *PBX1*, *MLLT3*, and *MEIS1* genes (Figures 5.9a and 5.10c), and high HOXA9 binding motif accessibility (Figure 5.9c), all related to self-renewing early haematopoietic cells. Mk/EP cells also showed chromatin profiles of highly primitive cells, but additionally had increased accessibility of genes like *HBB*, *GYPA*, and *GATA1* (Figures 5.9a and 5.10c). In contrast, the third and largest entity was composed from a continuum of lympho-myeloid progenitor-like states. Here, a relatively primitive cluster resembling LMPPs was detected (cluster: LMPP-like), in which open chromatin was enriched for immature lympho-myeloid signatures (Figures 5.9b and 5.9d). Three more differentiated progenitor-like clusters were also detected. One resembled progenitors of the monocytic/neutrophilic lineage (cluster: Mono/Neut progenitor-like), and was enriched for GMP, macrophage, monocyte, and neutrophil progenitor signatures (Figures 5.9b and 5.9d). The second resembled pDC and natural killer (NK) progenitors (cluster: pDC/NKP-like), where open chromatin was enriched in pDC precursor, monocyte/dendritic, and NK-cell signatures (Figures 5.9b and 5.9d). The third population resembled progenitors of the B-cell lineage (cluster: B progenitor-like) and, consistently, displayed chromatin characteristics typical

of more committed B-cell progenitors, like pre-B cells and CLPs (Figure 5.9b and 5.9d), high PAX5 binding motif accessibility (Figure 5.9c), and higher *CD79*, *PAX5*, and *EBF1* accessibility (Figures 5.9a and 5.10c). Finally, an intermediate lympho-myeloid progenitor population (cluster: LM progenitor-like), displaying chromatin characteristics of both myeloid and the lymphoid lineages (Figures 5.9a-b, 5.9d, and 5.10b), was detected. Given that immunophenotypic LMPP and GMP were expanded (Figure 5.2b), as frequently occurs in $CD34^+$ AML in which these populations harbour LSC activity [100], I inferred that the continuum of lympho-myeloid progenitor-like states represented the hierarchy of transformed leukaemic cells. Within the leukaemic hierarchy, each sub-population resembled distinct healthy lympho-myeloid, myeloid, and lymphoid progenitor counterparts. This was in line with malignant AML cells molecularly resembling normal progenitors [100] and being hierarchically organized [88, 251–254]. Projection onto the healthy reference further suggested these cells resembled early/intermediate progenitors of the myeloid and lymphoid lineages. In support of these progenitor-like populations representing transformed leukaemic cells was the fact that they did not form a continuum with HSC/MPP, as would be expected in a normal BM [80, 81, 88, 91, 398](Figure 5.10a). This suggested these transformed progenitor-like cell states had a chromatin landscape which differed substantially from normal progenitors, confirming that leukaemic cells exist in abnormal molecular states not observed in healthy haematopoiesis [252]. Hence, whilst HSC/MPP and Mk/EP cells likely represented either the residual normal or pre-leukaemic HSPCs, the lympho-myeloid entity comprised abnormal leukaemic cells. Consistently, *SETBP1*, typically expressed in LSCs [244], as well as *FLT3*, were most accessible in the leukaemic clusters (Figure 5.10c). Overall, scATAC-seq data identified distinct cellular states within the AML sample, which consisted of primitive HSC/MPP cells, mature T/NK cells, and leukaemic lympho-myeloid progenitor-like cells. Chromatin accessibility analysis within the leukaemic continuum further characterized the hierarchy of leukaemic differentiation and identified leukaemic sub-populations resembling normal progenitors.

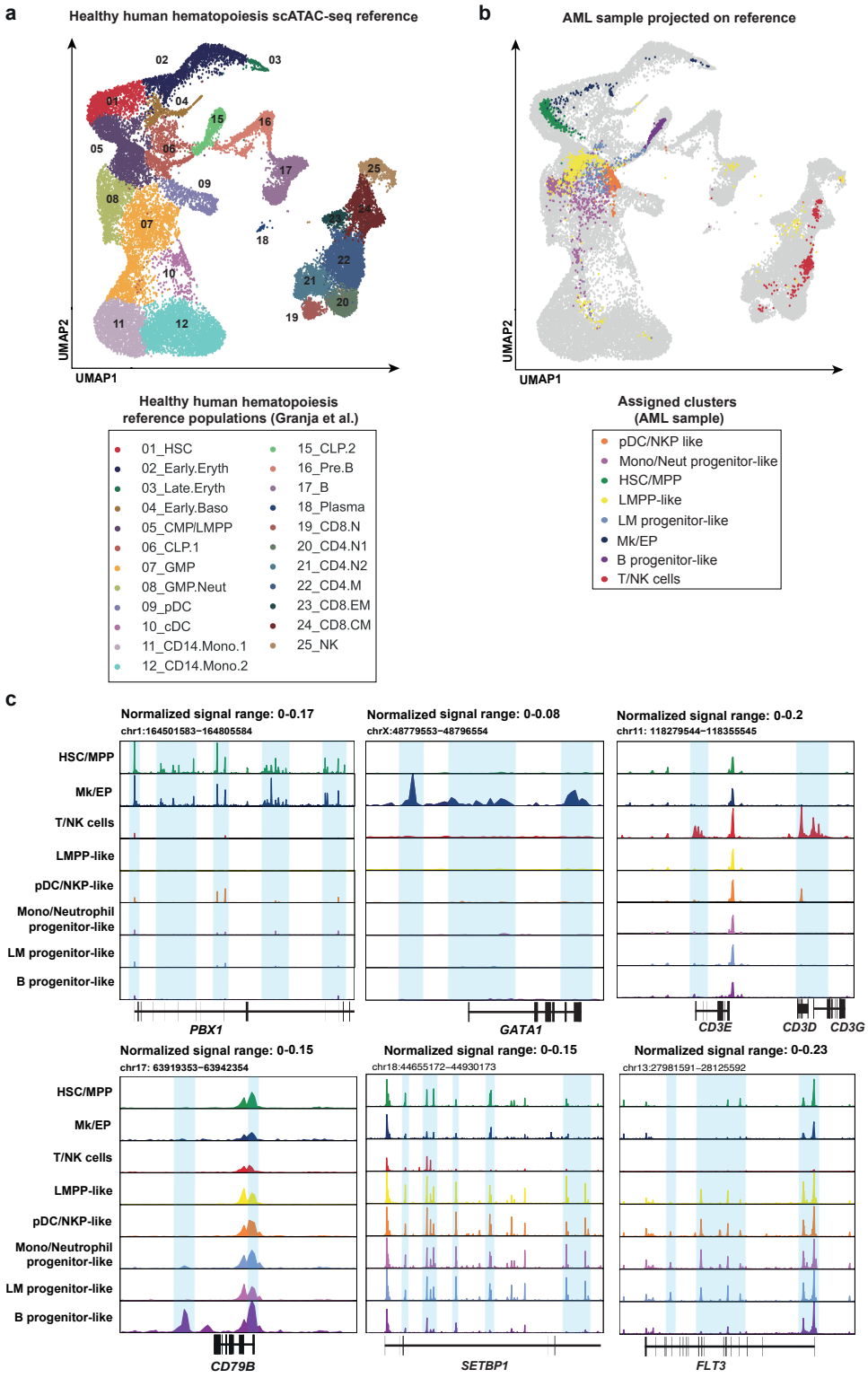


Figure 5.10: *(Legend on next page)*

Figure 5.10 ***(previous page)*****: Mapping to a scATAC-seq reference and cluster-specific chromatin accessibility. (a)** Left: Published scATAC-seq UMAP embedding providing a normal human haematopoiesis reference map [88]. Cluster labels (below) are retained from the original publication. **(b)** Projection of my AML dataset, labelled by assigned clusters (Figure 5.8), onto the reference embedding (grey) from (a). **(c)** Cluster-specific normalized scATAC-seq fragment coverage (y-axis) around the *PBX1*, *GATA1*, *CD3E/D/G*, *CD79B*, *SETBP1*, and *FLT3* genes.

5.3.3 Integration of clonal identity with chromatin accessibility

Having established the clonal (Figure 5.7) and cellular (Figure 5.8) hierarchies, I next investigated the correlation between the two. Mapping the genetic clonal identity onto chromatin accessibility for each single cell revealed a substantial segregation of different genetic clones into distinct cellular populations (Figure 5.11a). 97.5% of mature T/NK cells were devoid of driver somatic mutations (WT) and only 2.5% harboured a single *TET2* mutation (Figures 5.11a-b). This is consistent with T-cells being long-lived and for the most part not inheriting driver somatic mutations. Conversely, 87% of HSC/MPP and 82.5% of Mk/EP cells harboured a single *TET2* mutation. These cells were intermingled with a minority of *TET2/ASXL1*-double-mutant, as well as WT cells (Figures 5.11a-b). Hence, HSC/MPP and Mk/EP populations contained a mixture of genetically normal and pre-leukaemic cells. This was consistent with recent findings that dormant HSCs from AML patients are either WT or pre-leukaemic and do not harbour transforming mutations [252]. Meanwhile, the more evolved *RUNX1*-mutant and *RUNX1/BCOR*-mutant clones were fully segregated into the leukaemic lympho-myeloid continuum. In fact, $> 99\%$ of cells within the leukaemic hierarchy belonged to TAR/TARB1/TARB2 clones (Figures 5.11a-b). This suggested that the *RUNX1* mutation was the main genetic driver event associated with transformation to full-blown leukaemia. Of note, given that most cells were homozygous mutant for *RUNX1* (Figure 5.6a) and thus likely harboured a complete *RUNX1* LoF, it may not be surprising that a major disruption to the lympho-myeloid differentiation process occurred within mutant cells (section 1.1.3). Leukaemic cells belonging to the TAR clone intermingled with more evolved *BCOR*-mutant cells, suggesting

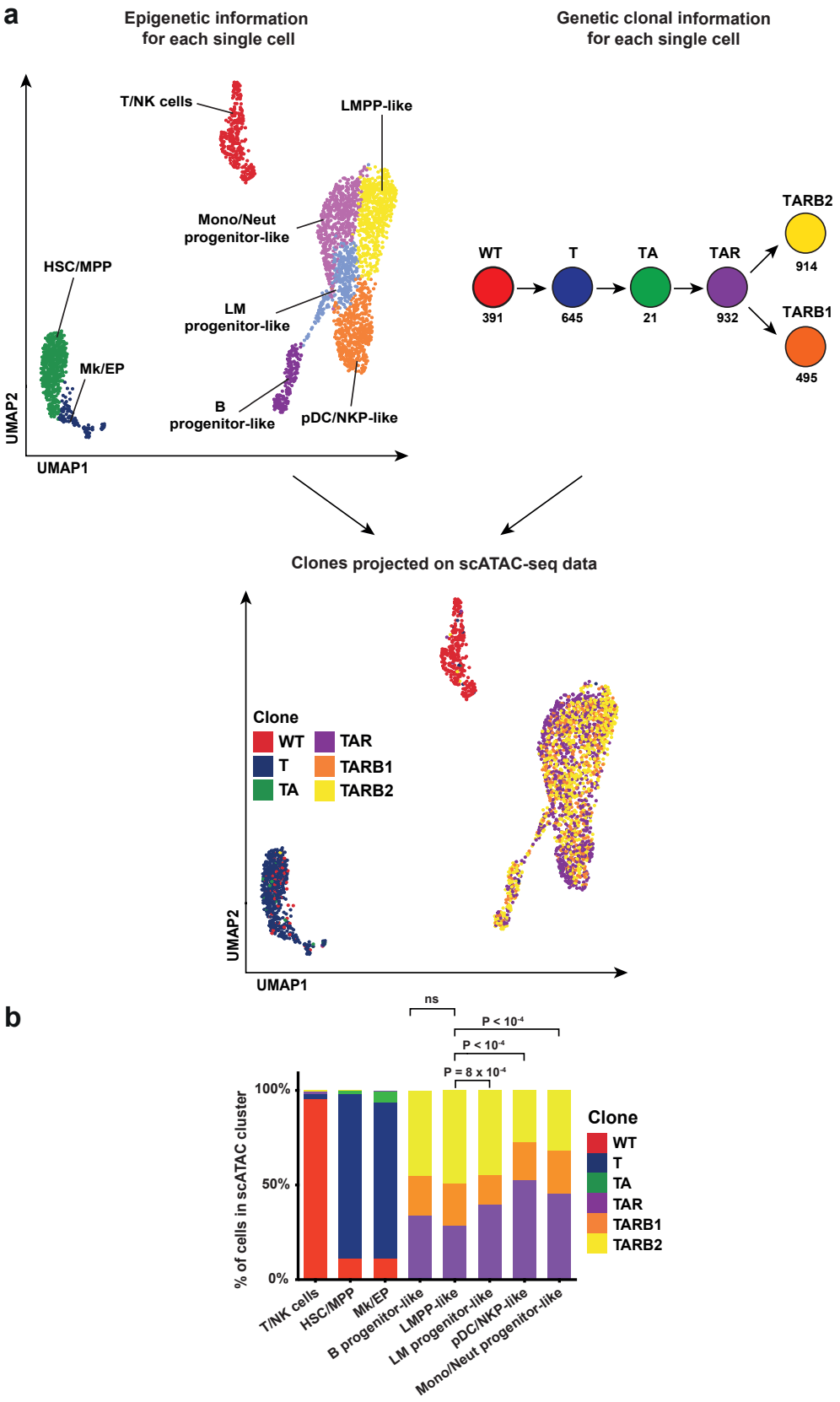


Figure 5.11: *(Legend on next page)*

Figure 5.11 ***(previous page)*****: Integration of genotypes with chromatin accessibility. (a)** Top left, UMAP from Figure 5.8; top right, clonal structure from Figure 5.7b. Bottom, UMAP from Figure 5.8 labelled by genetic clone (from Figure 5.7a). Cell numbers: WT, 377; T, 576; TA, 19; TAR, 861; TARB1, 446; TARB2, 834. **(b)** Clonal composition of each cell cluster. Differences in contribution of *BCOR*-mutant clones between the LMPP-like cluster and other leukaemic clusters were evaluated via the Fisher's exact test.

that further *BCOR* mutations did not form novel epigenomic states and that the effects of *BCOR* mutations were more subtle. Interestingly, *BCOR*-mutant cells clustered significantly more frequently within the least mature LMPP-like leukaemic population (Figure 5.11b).

Expectedly, TF binding motif accessibility (Figure 5.12a) mirrored clonal segregation into distinct stages of differentiation. WT cells were enriched for motifs bound by TFs important for T-cell biology, like LEF1 [399], EOMES [400, 401], and RUNX [402]. The pre-leukaemic T and TA clones, which segregated to HSC/MPP and Mk/EP, had high accessibility of motifs bound by TFs important for stem cell function (HOX, GATA2; section 1.1.1) and for megakaryocytic and erythroid differentiation (GATA1, NFE2/MAFF, and NFIA/NFIC family). In contrast, genetic leukaemic clones were enriched for motifs bound by lympho-myeloid TFs [116] (SPI1 and AP-1, CEBP, MEF2, IRF, and TCF family members; section 1.1.3). These analyses were further corroborated by TF binding motif footprinting (Figure 5.12b). Interestingly, although motif accessibility was highly similar in TAR, TARB1, and TARB2 clones, a certain degree of heterogeneity was observed. Specifically, motifs bound by MEF2C/D, oncogenic TFs associated with AML chemo-resistance and highly expressed in LSCs [244, 403–405], were strongly enriched in one of the *BCOR*-mutant clones (Figure 5.12a). Altogether, GTAC efficiently correlated genetic clonal hierarchy with distinct epigenomic cellular states. This identified the *RUNX1*-R201G mutation as the critical transforming genetic event, acting on a substrate of early pre-leukaemic mutations.

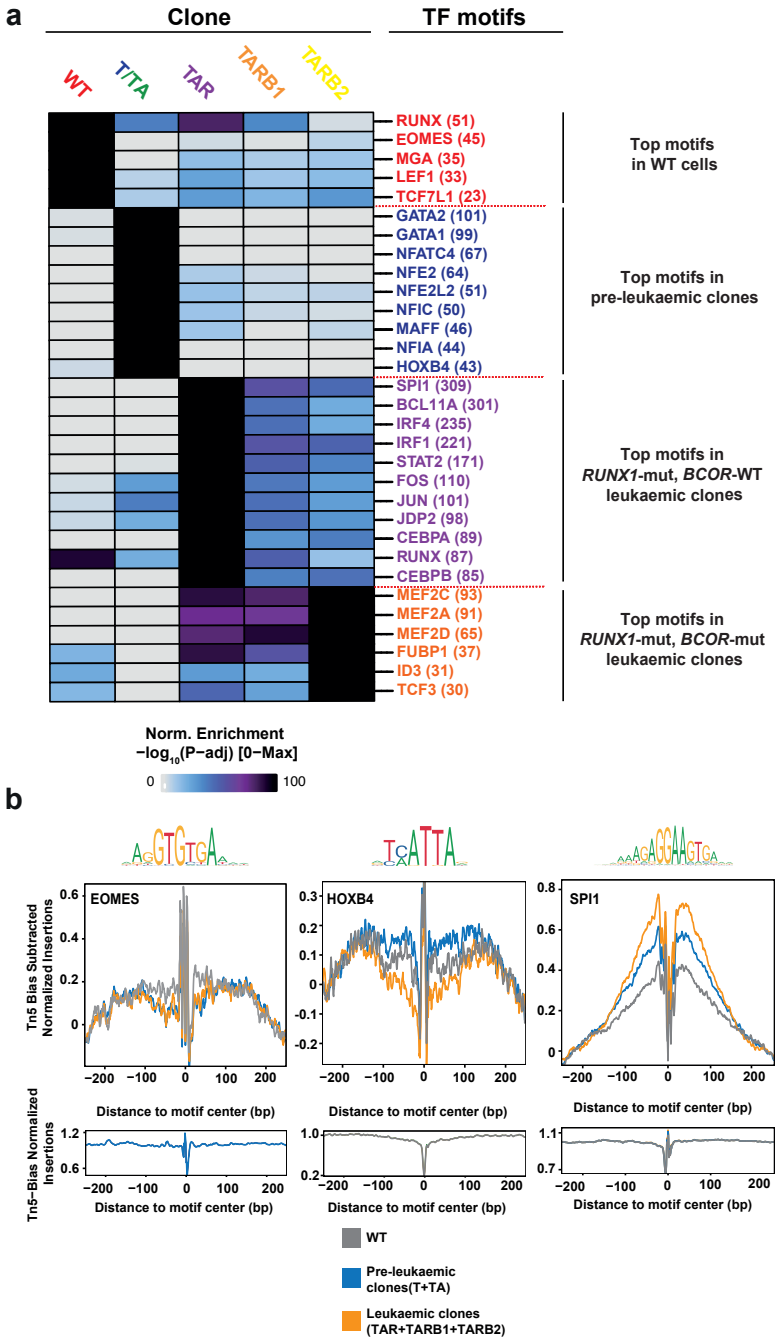


Figure 5.12: Clone-specific TF binding motif accessibility. **(a)** Enrichment of TF binding motifs in clone-specific marker peaks (FDR < 0.05 and $\log_2$FC > 0.5). Clones are represented as columns. TFs or TF families are indicated as rows. Heatmap colour scheme as in Figure 5.9d. Motifs for TFs expressed by haematopoietic cells [406, 407] are displayed. **(b)** TF footprints of motifs bound EOMES, HOXB4, and SPI1, comparing WT (n = 377 cells), pre-leukaemic (n = 595 cells), and leukaemic clones (n = 2,141 cells). Above, Tn5 bias-subtracted signal across the genomic area flanking the binding motif. Below, Tn5 binding bias across the same genomic area.

5.3.4 Epigenomic correlates of *RUNX1*-R201G-associated leukaemic transformation

Next, I wanted to characterize chromatin changes associated with *RUNX1*-R201G-linked leukaemic transformation. All *RUNX1*-mutant clones had increased accessibility of several genes which are more highly expressed in *RUNX1*-mutant versus *RUNX1*-WT AML patients [394, 408]. Exemplar genes were *IRF8*, *DNTT*, *BLNK*, and *MS4A1*, all of which lymphoid markers (Figure 5.13a). Higher accessibility of lymphoid markers in AML may seem counter-intuitive; nevertheless, there is a rarer subset of AML harbouring B/myeloid features, termed MPAL [409, 410]. MPAL is part of the group of leukaemias known as acute leukaemias of ambiguous lineage (ALAL). In adult B/myeloid MPAL, *RUNX1* mutation are enriched compared to T/myeloid MPAL [411]. Furthermore, double *RUNX1* mutations were noted in 4/10 of ALAL samples in another cohort [412]. Given that most leukaemic cells in my dataset had a homozygous *RUNX1* mutation (Figure 5.6a), and that a B progenitor-like leukaemic population was present (Figure 5.8), I speculate this patient had a B/myeloid MPAL. *MN1*, a potent AML oncogene [413] typically upregulated in *RUNX1*-mutant AML [408] and associated with poor prognosis [414], was also highly accessible in *RUNX1*-mutant clones (Figure 5.13a). Of note, MN1 can induce myeloid transformation by cooperating on chromatin with MEIS1, a well-characterized oncogenic TF, at shared binding sites [415]. Whilst open chromatin in pre-leukaemic clones was enriched in stemness genes and those regulated by PRC2 and MLL, *RUNX1*-mutant leukaemia was expectedly enriched in lympho-myeloid differentiation, active signalling, and immune activity (Figure 5.13b).

To identify putative transcriptional drivers of *RUNX1*-associated transformation, I directly compared TF binding motif accessibility between pre-leukaemic and leukaemic clones (Figure 5.13c). Expectedly, motifs bound by MECOM, and by HOX, GATA, and STAT family members were more open in pre-leukaemic clones. Interestingly, this held true for motifs bound by REL and NF KB motifs as well, consistent with mounting evidence about the role of this pathway in HSC [40, 79, 416]. In contrast, motifs bound by SPI1/B/C, CEBP family, and by TCF4,

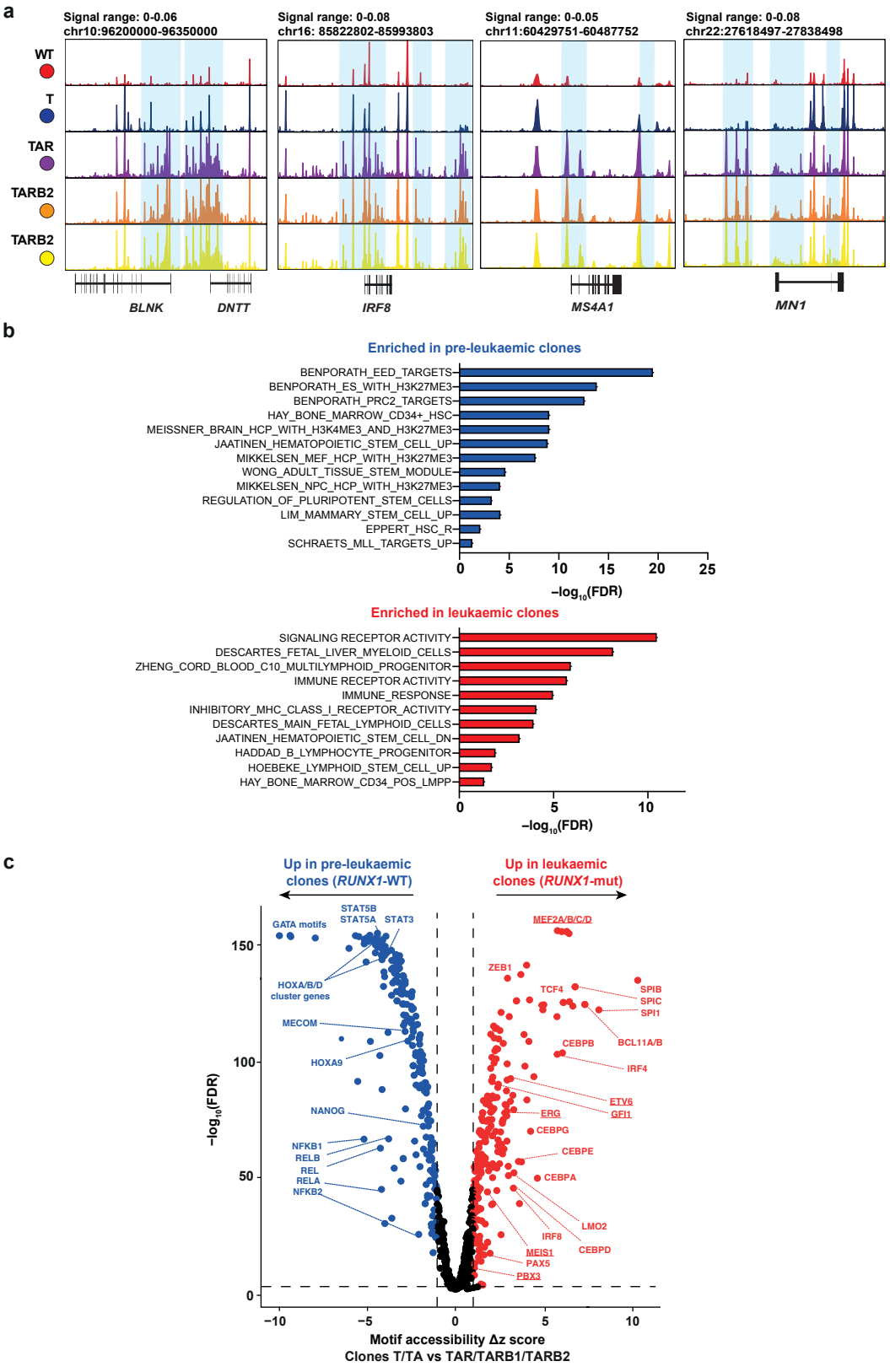


Figure 5.13: *(Legend on next page)*

Figure 5.13 *(previous page)*: Chromatin differences between pre-leukaemic and leukaemic genetic clones. (a) As in Figure 5.10c but comparing clone-specific chromatin accessibility around *BLNK*, *DNTT*, *IRF8*, *MS4A1*, and *MN1*. **(b)** Pathway enrichment of differentially accessible genes between pre-leukaemic and leukaemic clones. 625 top ranked genes (based on $\log_2$FC and showing FDR < 0.05) were used. **(c)** Differential TF binding motif accessibility between pre-leukaemic (n = 595 cells) and leukaemic (n = 2,141 cells) clones. Differentially accessible motifs (FDR < 0.05 and absolute $\Delta z > 1$; dashed lines) are labelled as red or blue dots. Underlined are motifs more accessible in leukaemic clones bound by those TFs reported to support HSC or LSC maintenance.

IRF4, ZEB1, and PAX5, associated with lympho-myeloid differentiation, were more accessible in leukaemic clones. However, despite leukaemic clones residing in a less stem-like and more progenitor-like state, they had higher accessibility of several motifs bound by TFs associated with either HSC or LSC maintenance, including PBX3, MEIS1, ERG, MEF2C/D, ETV6, and GFI1 motifs [24–26, 34, 417–420] (Figure 5.13c). These TFs may thus drive aberrant maintenance of *RUNX1*-mutant leukaemic progenitors.

Of note, an important limitation of the above analyses is that they do not discern which differences were associated with *RUNX1* mutation versus those arising due to distinct differentiation stages of pre-leukaemic and leukaemic clones (Figure 5.11). To identify chromatin changes linked to *RUNX1* mutation acquisition, it is necessary to compare *RUNX1*-WT pre-leukaemic and *RUNX1*-mutant leukaemic clones at a similar stage of differentiation. This was not possible using the original UMAP embedding (Figure 5.8) where *RUNX1*-mutant and WT cells were fully separated, likely due to pronounced chromatin reprogramming induced by transformation. To enable this comparison, I integrated clonal information to the projection of cells on healthy haematopoiesis (Figure 5.14a), where cells are forced to map to their most similar differentiation stage in the healthy BM counterpart. Firstly, this confirmed that most *RUNX1*-WT pre-leukaemic cells resided in early HSC/MPP compartments, whilst *RUNX1*-mutant leukaemic cells mapped mostly onto early and intermediate lympho-myeloid progenitors.

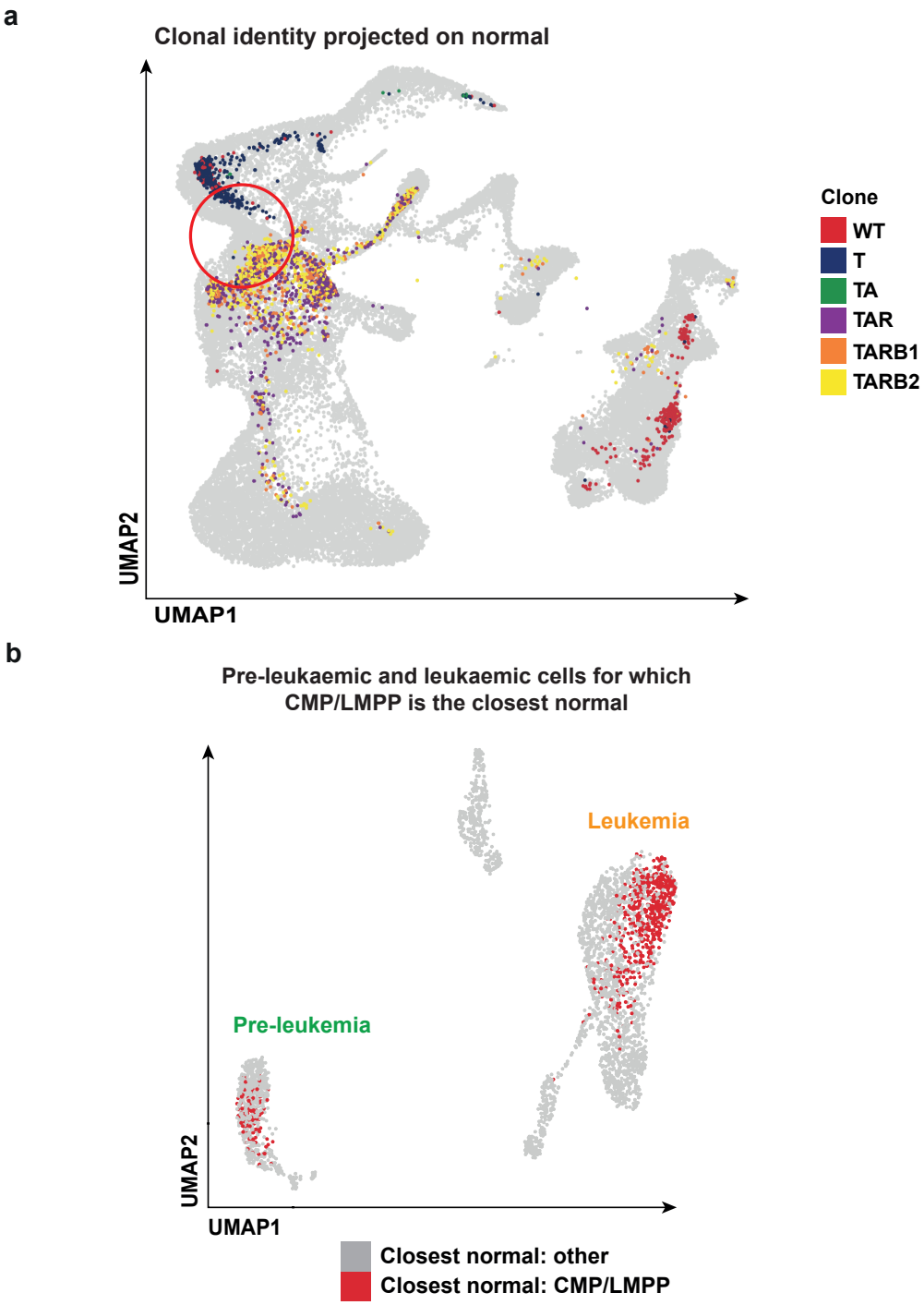


Figure 5.14: Differentiation stage at which leukaemic transformation occurred. **(a)** Same as in Figure 5.10b but labelled by genetic clonal identity of each cell. The red circle highlights the normal CMP/LMPP differentiation stage. **(b)** UMAP from Figure 5.8, in which cells mapping to normal CMP/LMPP are labelled in red.

However, both a subset of *RUNX1*-WT pre-leukaemic and *RUNX1*-mutant leukaemic cells mapped to the reference CMP/LMPP differentiation stage (Figure 5.14a, approximated by the red circle). Specifically, 97/595 *RUNX1*-WT (T or TA) cells, and 390/591 *RUNX1*-mutant (TAR/TARB1/TARB2) LMPP-like leukaemic cells, mapped to CMP/LMPP (Figure 5.14b). Conversely, only one *RUNX1*-mutant cell mapped to normal HSC. This provides two potential models for the COO of leukaemic clones (Figure 5.15): (a) a cell bearing *TET2*/*ASXL1* mutations transited through the CMP/LMPP stage when the *RUNX1* mutation occurred, or (b) a self-renewing *TET2*/*ASXL1*-mutant HSC acquired the *RUNX1* mutation, but uncontrolled proliferation occurred only when its progeny reached the LMPP stage. Crucially, both models indicate substantial epigenomic reprogramming within *RUNX1*-mutant LMPPs, suggesting the LMPP may provide the earliest cellular substrate for the oncogenic activity of *RUNX1* mutations.

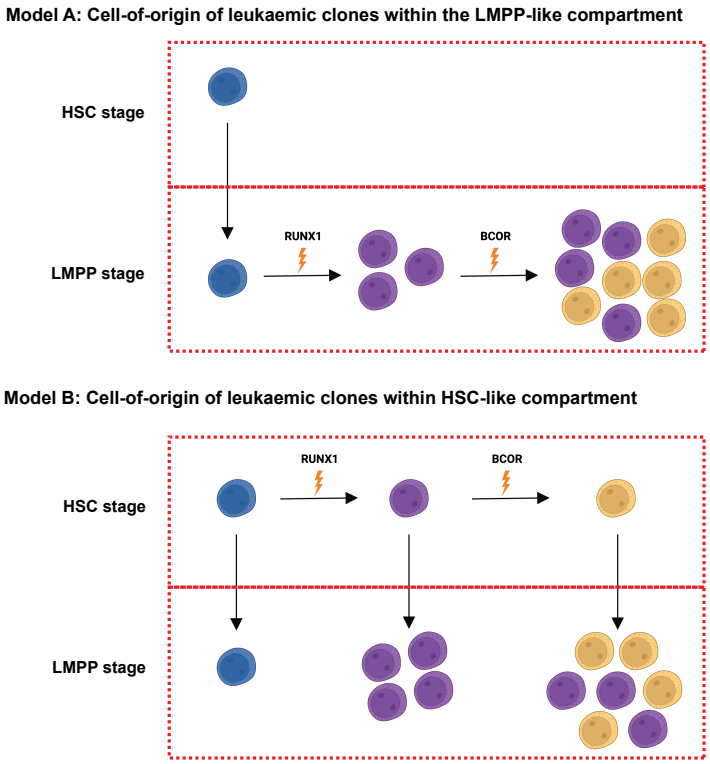


Figure 5.15: Two proposed models for the COO of the leukaemic clones Above, the *RUNX1* mutation was acquired at the LMPP stage and led to clonal advantage. Below, the *RUNX1* mutation was acquired at the HSC/MPP stage but conferred clonal advantage only at the LMPP stage.

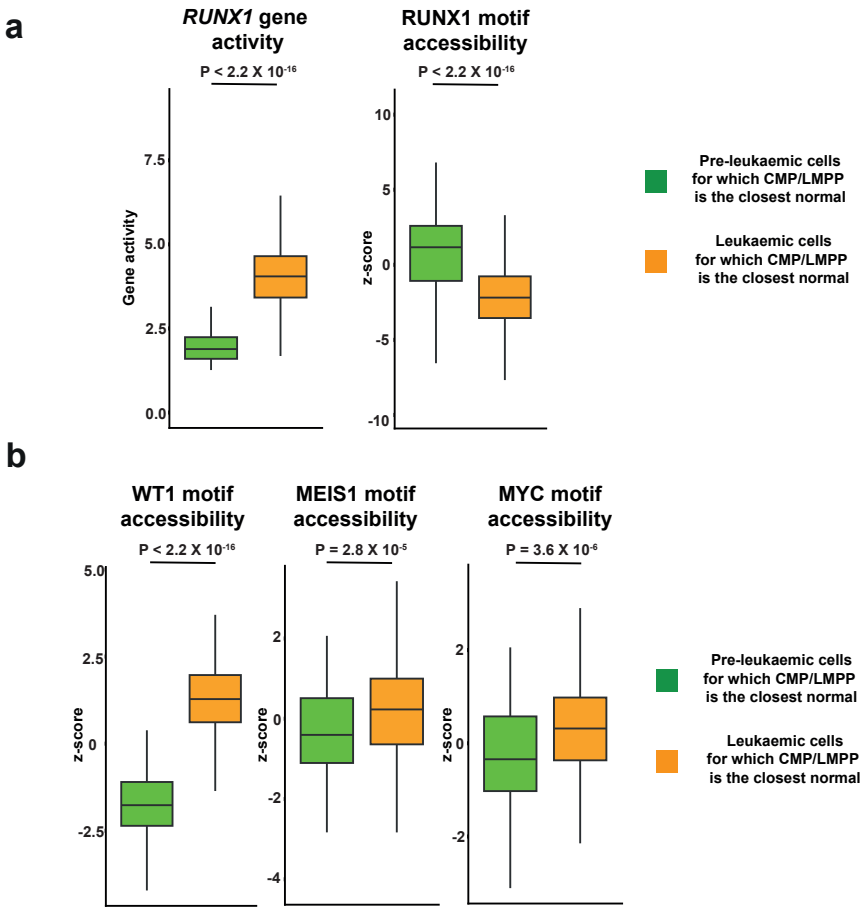


Figure 5.16: Selected chromatin differences between pre-leukaemic and leukaemic LMPPs. (a) Left, *RUNX1* gene accessibility in pre-leukaemic (n = 97) and leukaemic (n = 492) cells mapping to normal CMP/LMPP. *P*-value computed with the Wilcoxon rank sum test. Right, same panel structure but for the chromVAR-computed RUNX1 binding motif accessibility. **(b)** Same as the right plot in (a) but for WT1, MEIS1, and MYC motifs.

The above analysis identified *RUNX1*-mutant and *RUNX1*-WT cells residing within a similar differentiation stage (LMPP). To elucidate chromatin changes associated with the *RUNX1* mutation, I compared chromatin profiles of pre-leukaemic *RUNX1*-WT and leukaemic *RUNX1*-mutant cells mapping exclusively to CMP/LMPP. Interestingly, *RUNX1* was more accessible in leukaemic CMP/LMPP, suggesting high requirement for RUNX1 activity in these cells (Figure 5.16a, left). However, the RUNX1 binding motif was less accessible (Figure 5.16a, right), consistent with impaired RUNX1 DNA binding due to the R201G mutation. *RUNX1*-mutant CMP/LMPP had higher accessibility of motifs bound by WT1

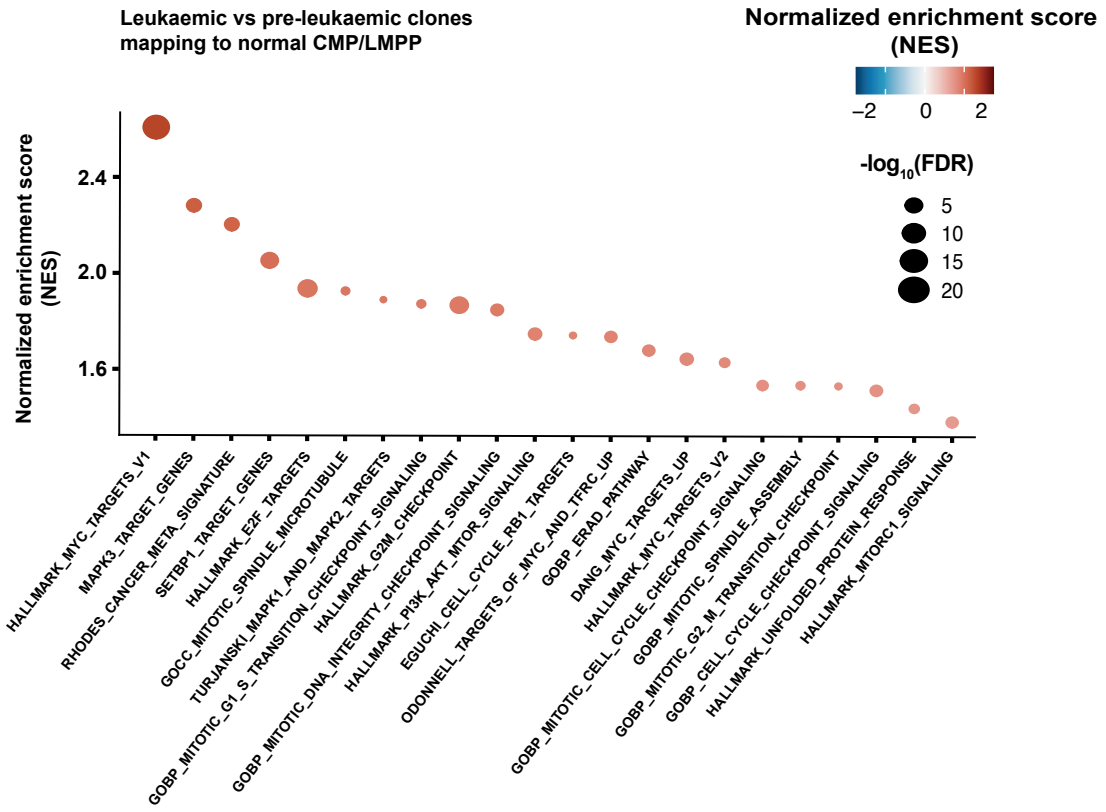


Figure 5.17: Leukaemic LMPPs display a pro-proliferative chromatin signature. Enrichment of genes more accessible in leukaemic versus pre-leukaemic cells mapping to normal CMP/LMPP in selected pathways (normalized enrichment score, NES). Dot sizes indicate -$\log_{10}$(FDR).

and MEIS1 (Figure 5.16b), oncogenic TFs supporting AML maintenance [420, 421]. This further supports that MEIS1 is likely to have a role in transformation of these *RUNX1*-mutant cells. Moreover, the motif bound by MYC was also more accessible in *RUNX1*-mutant CMP/LMPP (Figure 5.16b), suggesting an increased proliferative potential of these cells. Indeed, accessible chromatin in *RUNX1*-mutant CMP/LMPP was enriched for a series of cell cycle and stress response pathways, as well as those linked to cell growth, MAPK, and PI3K signalling (Figure 5.17). Several genes more accessible in *RUNX1*-mutant CMP/LMPP were genes commonly upregulated across various cancers, suggested by the '*Rhodes_Cancer_Meta_Signature*'.

These data indicate that leukaemic transformation of the LMPP, linked to the acquisition of the *RUNX1*-R201G mutation, correlates with an increasingly proliferative chromatin landscape. Whilst it is possible that loss of RUNX1 function

promotes some of these changes directly, the most parsimonious explanation is that *RUNX1*-deficient lympho-myeloid progenitors are arrested within a transit-amplifying population with high proliferative potential.

5.4 Consequences of subclonal *BCOR* mutations on leukaemic differentiation

BCOR mutations are recurrent in myeloid malignancy [422, 423]. In AML, *BCOR* mutations, predominantly LoF, co-occur frequently with those in *RUNX1*, suggesting mutation cooperativity [296, 297, 423]. Both murine and recent human studies suggested that *BCOR* loss dysregulates myeloid differentiation, which has been attributed to upregulation of stemness-associated gene programs [319, 424, 425]. However, the mechanistic link between *BCOR* mutations and increased stemness, as well as the molecular basis for the cooperativity between *RUNX1* and *BCOR* mutations, are not fully understood. *BCOR*, located on the X chromosome, encodes for a key component of the non-canonical Polycomb repressive complex 1.1 (PRC1.1). PRC1 complexes are repressive epigenetic regulators, which place a ubiquitin mark on the Lysine 119 residue of the histone H2A (H2AK119Ub1), thus promoting gene silencing [426–428]. Within the PRC1.1, BCOR associates with the E3-ubiquitin ligase RING1A/B, PCGF1, and RYBP/YAF2 proteins. The PUFD domain at the BCOR C-terminus is important for binding of BCOR to PCGF1 and consequent PRC1.1 assembly [429]. BCOR or other PRC1.1 components do not bind to chromatin directly and are instead dependent on KDM2B for chromatin recruitment [430, 431]. Accordingly, *BCOR* mutations unlink the PRC1.1 enzymatic complex from the chromatin-binding complex, leading to de-repression of PRC1.1 targets [432]. Although PRC1 often works in coordination with PRC2, it has also been shown that PRC1.1 binds a unique set of genes independently of PRC2, in human myeloid leukaemia cells [367].

Consistent with cooperation between *RUNX1* and *BCOR* mutations, in the patient studied above, two independent *BCOR* mutations arose subclonally to the *RUNX1* mutation. Both *BCOR*-mutant subclones reached a substantial size,

suggesting the acquisition of *BCOR* mutations conferred an additional fitness advantage on top of the *RUNX1* mutation. As the patient was male, both *BCOR* mutations were hemizygous (Figure 5.6). Both mutations occurred in exon 4 and were predicted LoF: one results in a premature stop codon, while the other is a frameshift mutation. Given that both are predicted to result in the loss of the PUFD domain, *BCOR*-mutant clones likely harbour a complete loss of PRC1.1 function, due to the inability of *BCOR* to bind PCGF1.

Hence, GTAC uniquely enabled to disentangle how subclonal *BCOR* mutations cooperate with a *RUNX1* mutation to alter gene regulation. Specifically, I asked:

1. How do *BCOR* mutations influence differentiation of leukaemic cells?

2. What is the epigenetic basis for the accumulation of *BCOR* mutant clones within the LMPP-like leukaemic population?

3. How to *BCOR* and *RUNX1* mutations cooperate to alter gene regulation?

5.4.1 Clonal basis for the LSC-like chromatin state within the leukaemic hierarchy

Although *BCOR*-mutant subclones (TARB1/TARB2) were intermingled with *BCOR*-WT, *RUNX1*-mutant (TAR) cells, TARB1/TARB2 cells accumulated within the LMPP-like population (Figure 5.11b). This observation prompted a detailed characterization of the LMPP-like cluster. Based on chromatin accessibility (Figures 5.9 and 5.10), I hypothesized that the LMPP-like cluster represented the least differentiated leukaemic population in this sample. To confirm this, I obtained a published LSC gene score [244], derived by comparing immunophenotypic LSCs ($Lin^-CD34^+CD38^-CD90^-$), leukaemic progenitors ($Lin^-CD34^+CD38^+$) and more differentiated leukaemic blasts (Lin^-CD34^-) in human AML. Indeed, accessible chromatin within LMPP-like cells was enriched for the LSC score (Figure 5.18a). Expectedly, chromatin in the genetically pre-leukaemic HSC/MPP population was also enriched for the LSC score, likely reflecting the fact that these genes are highly expressed in both LSC and HSC.

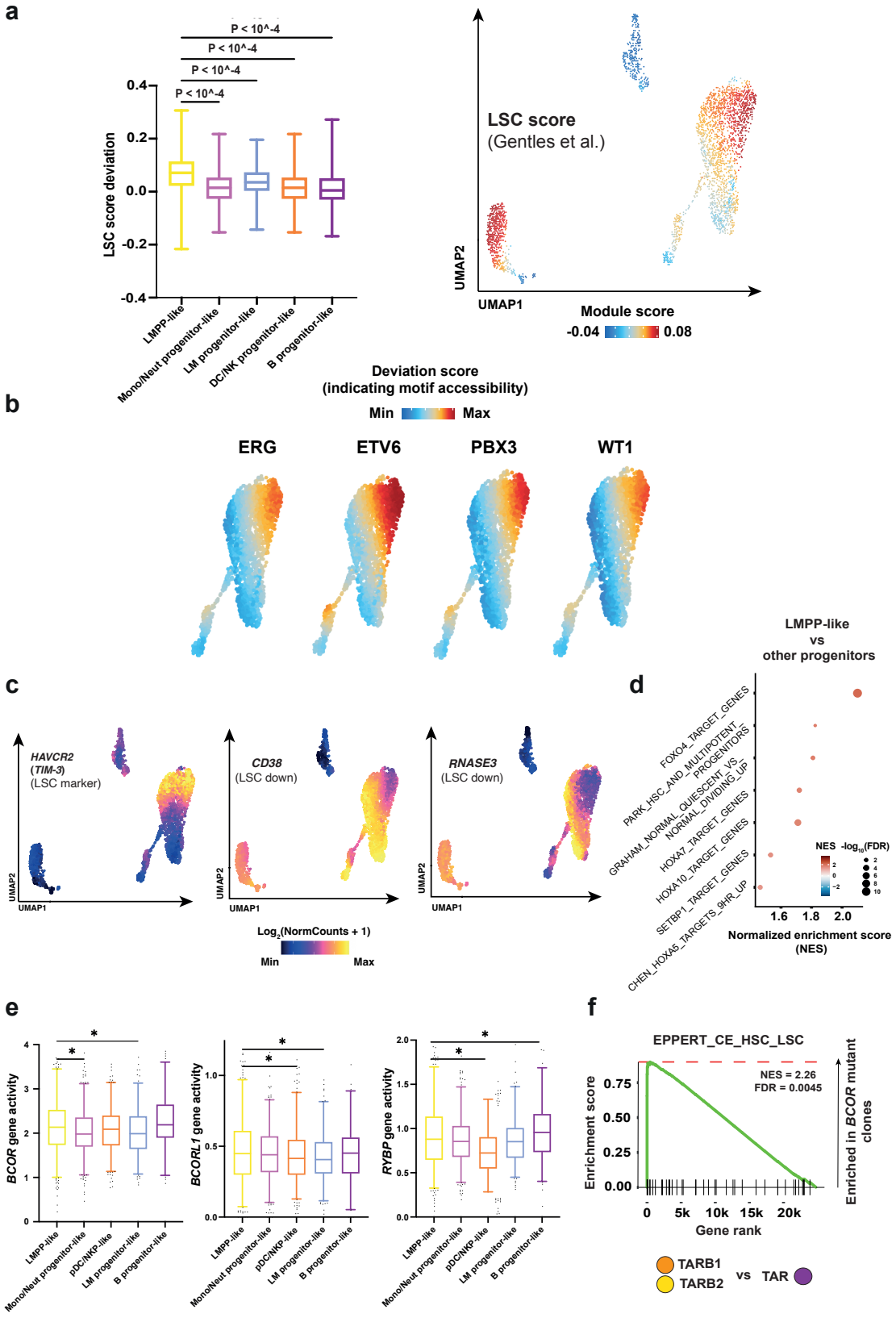


Figure 5.18: *(Legend on next page)*

Figure 5.18 *(previous page)*: Epigenetic resemblance of the LMPP-like cluster to LSCs. (a) Left, LSC module score [244] deviation (y-axis) across leukaemic progenitor clusters (x-axis). Median values are shown. Whiskers extend from the minimum to maximum value. *P*-values computed with the Kruskall-Wallis test with Dunnett's multiple comparisons. Right, the UMAP overlaid with LSC module score. **(b)** Leukaemic clusters overlaid with ERG, ETV6, PBX3, and WT1 motif accessibility. **(c)** UMAP overlaid with gene activity scores for *HAVCR2*, *CD38*, and *RNASE3*. **(d)** Selected pathways enriched in chromatin more accessible in LMPP-like cells compared to all other leukaemic clusters. **(e)** Accessibility of genes encoding selected PRC1.1 members across leukaemic clusters. *P*-values computed using the Kruskall-Wallis test with Dunnett's multiple comparisons. A star indicates significant differences. **(f)** Enrichment of genes more accessible in TARB1/TARB2 versus TAR cells in genes upregulated in HSCs and LSCs compared to normal and leukaemic progenitors [231].

However, since the HSC/MPP population was either WT or genetically pre-leukaemic, these data suggested that the LMPP-like cluster was likely the most LSC-enriched cell fraction in this sample. Thus, I refer to these cells as molecularly defined LSCs. Consistently, the LMPP-like cluster displayed higher accessibility of ERG, ETV6, PBX3, and WT1 binding motifs, important for LSC function (Figure 5.18b). The *TIM3* gene, an LSC marker (section 1.2.3, was also highly accessible in LMPP-like cells (Figure 5.18c). Conversely, *CD38* was less accessible in the LMPP-like cluster. Moreover, *RNASE3*, downregulated in LSCs [244], was also less accessible in LMPP-like cells (Figure 5.18c). Accordingly, when compared to other leukaemic progenitors, accessible chromatin in LMPP-like cells was enriched for FOXO and HOXA targets, linked to stemness and quiescence (section 1.1.1, as well as for targets of SETBP1, highly expressed in LSCs [244, 433] (Figure 5.18d). Accordingly, the chromatin landscape of LMPP-like cells resembled that of quiescent and multipotent cells (Figure 5.18d). Altogether, the LMPP-like cluster displayed an LSC-like chromatin milieu. It is notable that, compared to other leukaemic clusters, the LMPP-like population displayed higher accessibility of *BCOR* and other selected PRC1.1 components, like *RYBP* and *BCORL1* (Figure 5.18e). This suggests a requirement for PRC1.1 activity in LMPP-like cells, possibly to differentiate from this stage, in line with *BCOR*-deficient cells accumulating within the LMPP-like cluster. Compared to TAR leukaemic counterparts, TARB1/TARB2 cells were

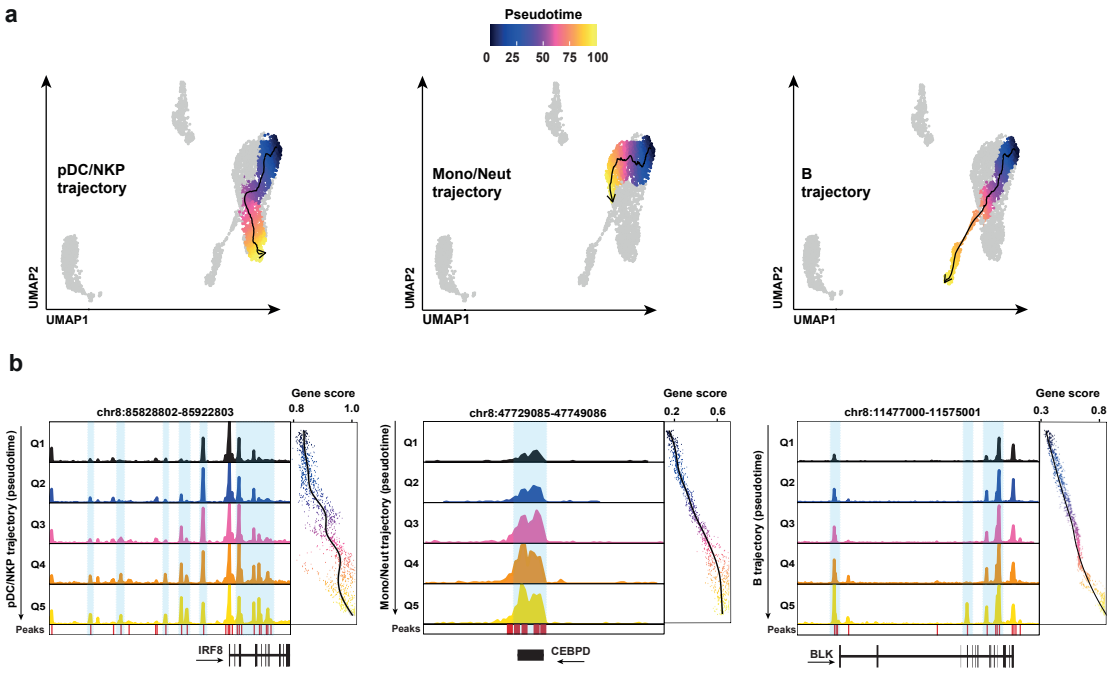


Figure 5.19: Differentiation trajectories within the leukaemic hierarchy. (a) Three differentiation pathways undertaken by leukaemic cells. UMAPs show pseudotime estimation for the three lineages. Cells are colour-coded by pseudotime score. **(b)** Variation in chromatin accessibility around *IRF8*, *CEBPE*, and *BLK* genes, as cells progress through the pDC/NKP, Mono/Neut, and B trajectory, respectively. Each trajectory was segregated into five quantiles from low to high pseudotime. On the right of each graph, gene accessibility for the respective genes is quantified across pseudotime.

enriched for genes upregulated in HSCs/LSCs (Figure 5.18f). Overall, *RUNX1*-mutated leukaemic cells which acquire *BCOR* LoF mutations are enriched in an LSC-like epigenetic state.

5.4.2 Differentiation delay of *BCOR*-mutant leukaemic cells

The data above suggested *BCOR*-mutant subclones differentiated less efficiently compared to their *BCOR*-WT counterparts. To characterize differentiation dynamics of leukaemic clones, I computed 3 possible differentiation trajectories within the leukaemic hierarchy (Figure 5.19a). All three trajectories depart from the LMPP-like compartment and extend along the pDC/NKP, Mono/Neut, or the B cell lineages (Figure 5.19a).

Of note, while the pDC/NKP and B trajectories transited through the intermediate LM progenitor-like state, the Mono/Neut trajectory extended directly from

LMPP-like towards Mono/Neut progenitor-like cells, resembling GMPs, which are directly downstream of LMPPs. Accordingly, *IRF8*, *CEBPD*, and *BLK* genes, important for pDCs, monocytes/neutrophils, and B-cells, respectively, gradually gained accessibility along the respective trajectories (Figure 5.19b). Hence, the leukaemic hierarchy partially mimicked normal differentiation pathways. Consistent with inefficient differentiation of *BCOR*-mutant cells, I observed a statistically significant accumulation of *BCOR*-mutant cells earlier in pseudotime (Figures 5.20a-b) for all differentiation pathways. Interestingly, the distribution of TARB1 and TARB2 clones across pseudotime was different (Figure 5.20a). In fact, TARB2 (the larger of the two) was more delayed across the Mono/Neut and pDC/NKP trajectories, but not across the B-cell trajectory (Figure 5.20a).

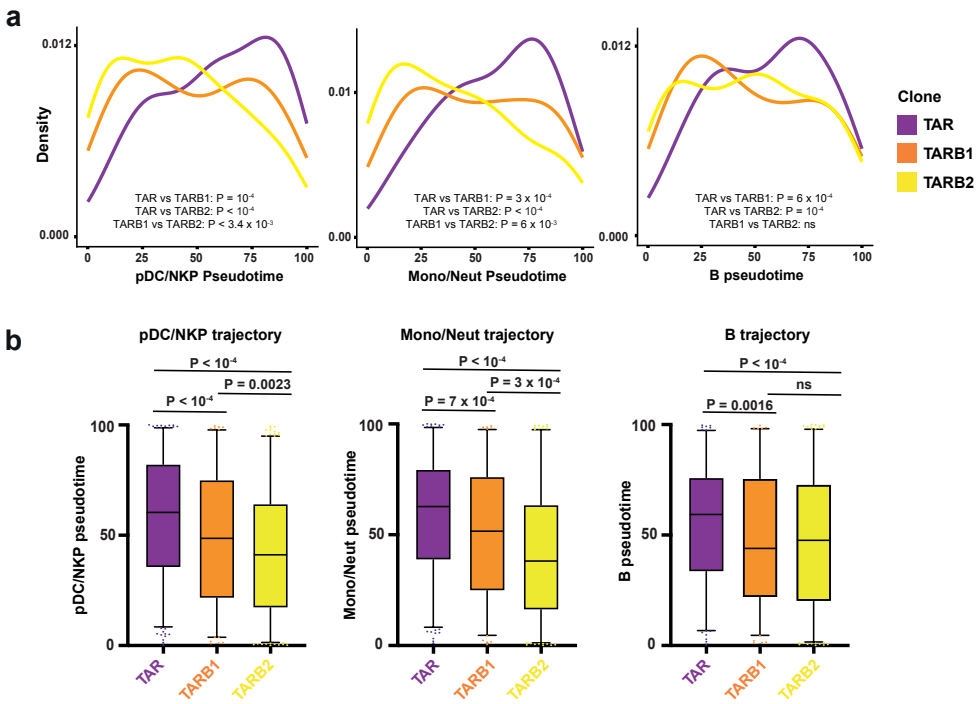


Figure 5.20: Clonal frequency across the leukaemic hierarchy. (a) For each differentiation trajectory, clonal density is plotted across pseudotime. The Kolmogorov-Smirnov test was used to identify significant differences in density distributions. *P*-values indicated for each comparison. **(b)** Pseudotime scores for TAR (purple), TARB1 (orange), and TARB2 (yellow) cells across the trajectories. *P*-values computed via the Wilcoxon rank sum test.

This suggested: (a) *BCOR*-mutant leukaemic cells differentiated less efficiently compared to *BCOR*-WT leukaemic cells; and (b) there are shared, but also distinct

differentiation properties of the two *BCOR*-mutant clones. The accumulation of *BCOR*-mutant clones within the immature LMPP-like cluster was reflected in the delayed accessibility of motifs bound by differentiation-associated TFs TCF4 and CEBPA in trajectories containing exclusively *BCOR*-mutant cells (Figures 5.21a-b). In other words, *BCOR*-mutant cells accumulated within a state with low activity of these lineage-related TFs (section 1.1.3).

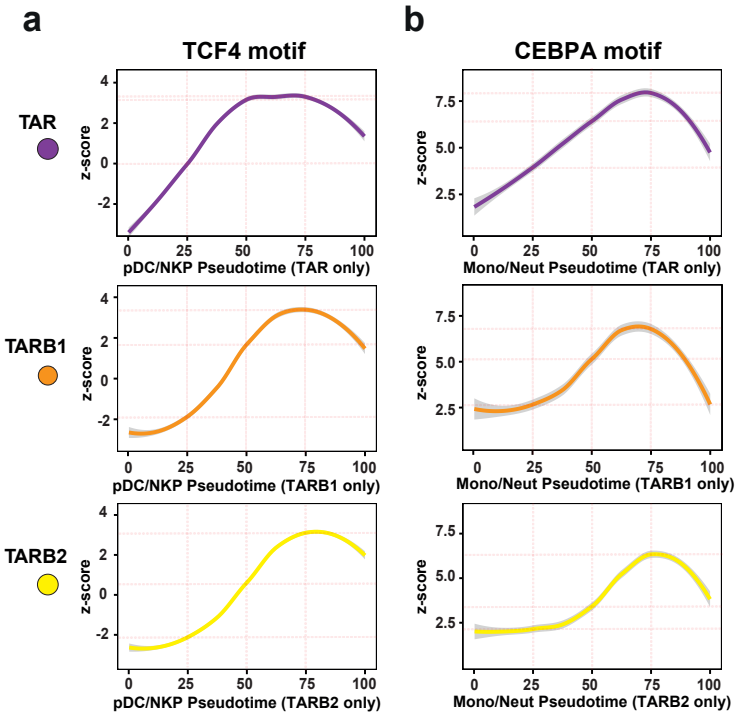


Figure 5.21: Clone-specific TF binding motif accessibility through pseudotime. **(a)** TCF4 motif accessibility across pDC/NKP pseudotime within clone-specific trajectories. **(b)** CEBPA motif accessibility across Mono/Neut pseudotime within clone-specific trajectories.

Consistently, motifs bound by TFs mediating AML maintenance and self-renewal (ERG, PBX3, FLI1, ETV6, YY1, NF-Y family) [24–26, 34, 43, 417–419], which showed highest accessibility in LMPP-like cells, lost accessibility in a slower fashion in *BCOR*-mutant cells (Figure 5.22). Indeed, the TARB2-specific pDC/NKP differentiation trajectory contained more cells displaying intermediate accessibility of these stemness-associated TFs (light orange on the heatmap), in line with a less efficient exit of *BCOR*-mutant cells from an LSC-like state.

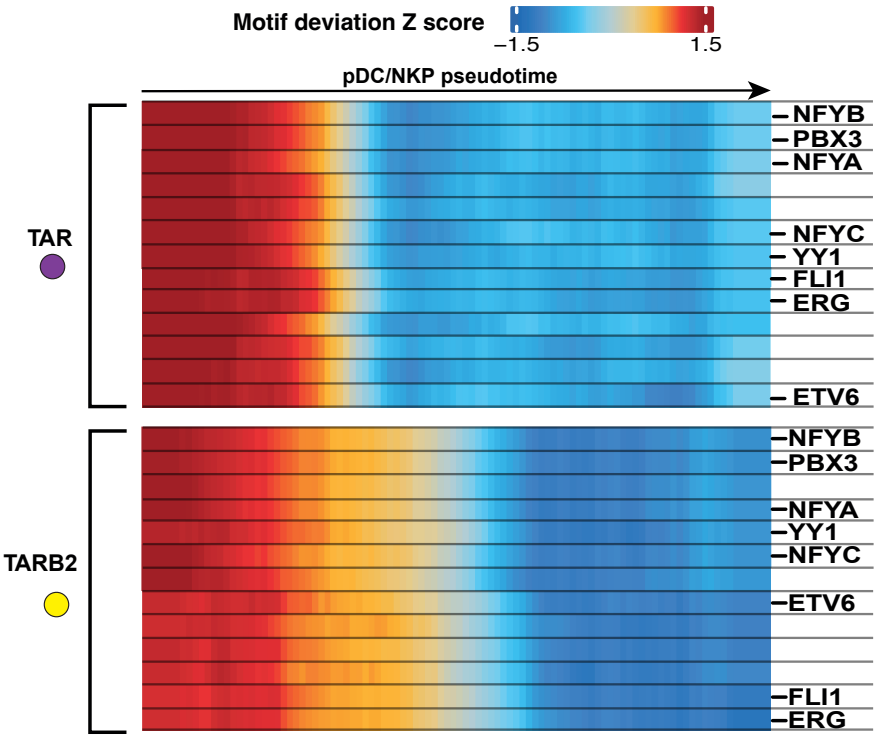


Figure 5.22: Unbiased analysis of TF motif accessibility through pseudotime. Heatmaps showing the variation of motif accessibility across the TAR-specific (above) and TARB2-specific (below) pDC/NKP trajectory (x-axis), for a set of most variable motifs across pseudotime (rows). Motifs bound by TFs functionally shown to promote self-renewal and/or AML maintenance are labelled

.

In contrast, the TAR-specific trajectory contained less cells in this 'intermediate state', suggesting quicker silencing of stemness programs (Figure 5.22). Of note, several of these motifs (for example, those bound by FLI1/ERG/ETV6) share a highly similar DNA sequence and thus co-vary in an almost identical manner, precluding robust conclusions on which of these TFs is functionally relevant for the LSC-like state in this context. In line with inefficient myeloid differentiation, there was delayed accessibility of regulatory elements co-accessible with *TLR4*, and of peaks mapping to *CEBPD*, as TARB2 leukaemic cells differentiated (Figure 5.23a). Accordingly, compared to TAR, TARB2 Mono/Neut progenitor-like cells displayed lower accessibility of genomic peaks proximal to genes associated with neutrophilic and monocytic differentiation (Figure 5.23b). Altogether, *BCOR* mutations disrupt myeloid differentiation and delay leukaemic cells at an LSC-like state.

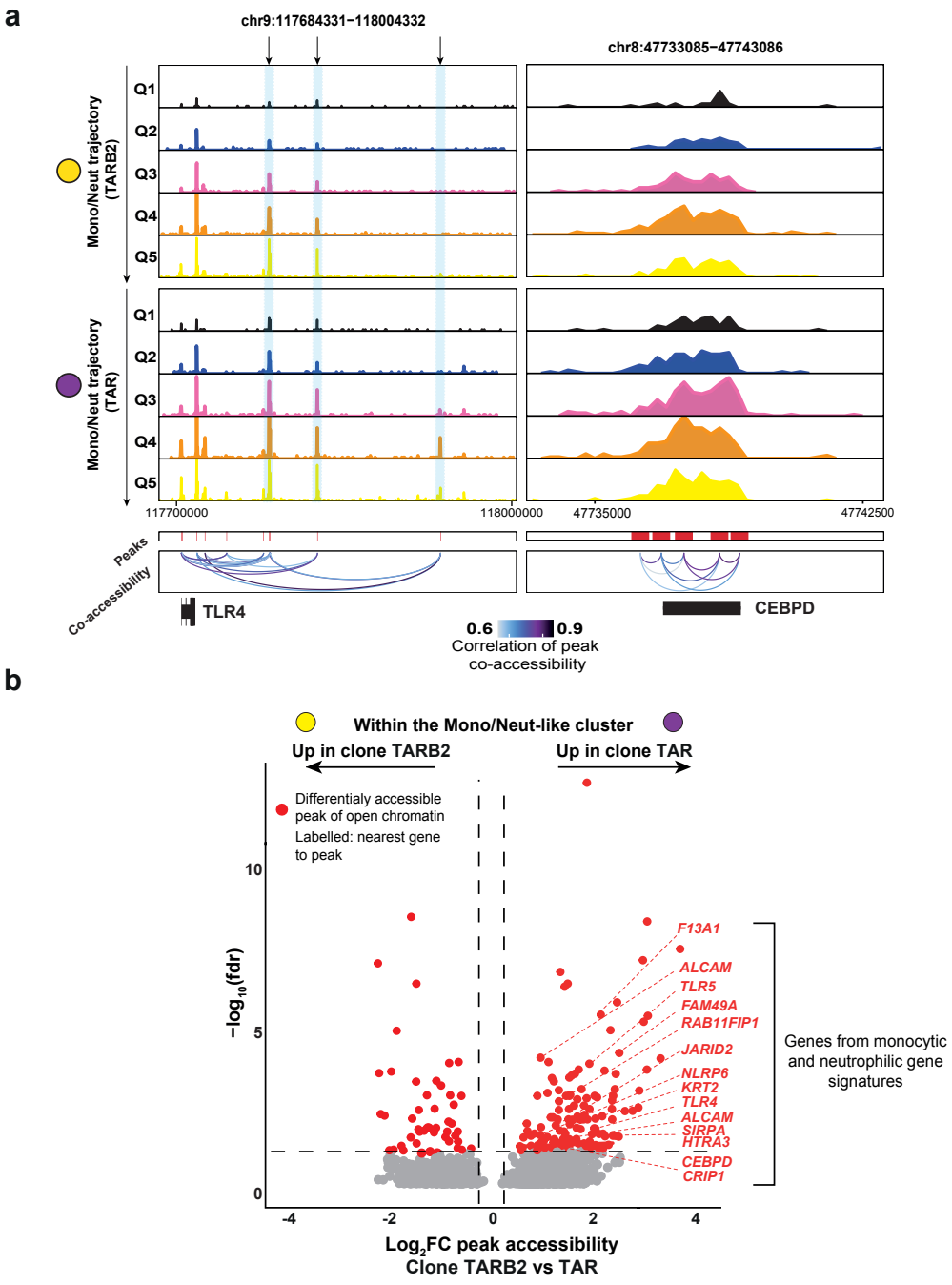


Figure 5.23: Clone-specific chromatin changes associated with myeloid differentiation. **(a)** Variation in chromatin accessibility around *TLR4* (left) and *CEBPD* (right) genes, across the TARB2-specific (above) and TAR-specific (below) Mono/Neut trajectory, each split into five quantiles (Q1-Q5). Peaks co-accessibility indicated below. **(b)** Differentially accessible peaks of open chromatin (FDR < 0.05 and $\log_2$FC > 0.25; dashed lines) between TARB2 (n=293 cells) and TAR (n=167 cells) Mono/Neut progenitor-like cells. The nearest gene is labelled for specific peaks. Labelled genes belong to monocytic/neutrophilic progenitor signatures.

5.5 Putative epigenetic mechanisms underlying selection of *BCOR*-mutant clones

I next investigated the mechanistic basis for the retention of *BCOR*-mutant cells within an LSC-like molecular state. Moreover, I asked whether altered gene regulation in *BCOR*-mutant LMPP-like cells could provide insights into *BCOR* and *RUNX1* mutation cooperativity.

5.5.1 Epigenetic consequences of *BCOR* mutations within the LMPP-like cluster

To understand the epigenetic basis for clonal advantage of *BCOR*-mutant cells within molecularly defined LSCs (LMPP-like cells), I compared gene accessibility between TAR and TARB (TARB1 or TARB2) cells specifically within the LMPP-like cluster (Figures 5.24a-b). Moreover, to distinguish between likely direct and indirect effects of *BCOR* mutations, I compiled a list of genes most frequently bound by PRC1.1 (i.e. bound by KDM2B) in 6 primary human AML samples (indicated as green dots in Figures 5.24a-b), using previously published ChIP-seq data [367].

Whilst a subset of putative PRC1.1 direct targets were more accessible in *BCOR*-mutant clones, consistent with de-repression of these genes upon PRC1.1 loss, some were surprisingly more accessible in *BCOR*-WT LMPP-like cells. Three putative PRC1.1 direct targets were more accessible in both TARB1 and TARB2 LMPP-like cells: *HIVEP2*, *INPP4B*, and *SOCS2*. Whilst the function of *HIVEP2* has not been documented in haematopoiesis, both *INPP4B* and *SOCS2* were previously implicated in HSC and LSC biology. Notably, *INPP4B* and *SOCS2* are more highly expressed in HSCs/LSCs compared to more differentiated normal/leukaemic progenitors [231, 434–436], suggesting a role in self-renewal and quiescence of primitive BM cells. Both genes are also associated with adverse prognosis and early relapse in myeloid leukaemia [435, 437–439]. SOCS2 is a negative regulator of JAK/STAT signalling and protects HSCs from exhaustion under stress conditions [435]. Moreover, high expression of *SOCS2* is a strong predictor of poor patient

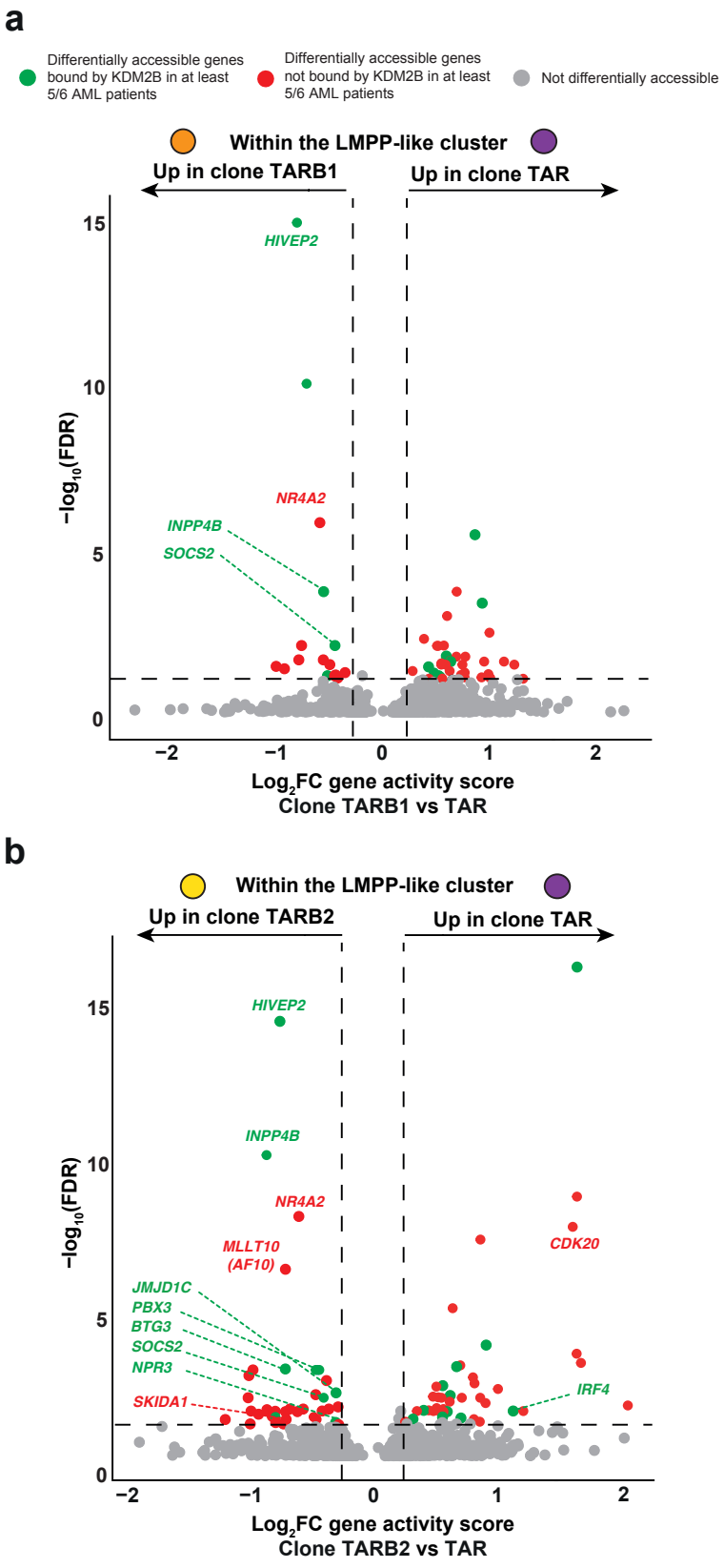


Figure 5.24: *Legend on next page*

Figure 5.24 *(previous page)*: Differentially accessible genes between *BCOR*-mutant and *BCOR*-WT LMPP-like cells. (a) Differentially accessible genes between TARB1 (n=130 cells) and TAR (n=167 cells) LMPP-like cells (FDR < 0.05 and absolute $\log_2$FC > 0.25; dashed lines). Differentially accessible genes bound by KDM2B in at least 5/6 AML patients (ref) are labelled in green. **(b)** As in (a) but comparing TARB2 (n=293 cells) and TAR (n=167 cells) clones.

survival, and *SOCS2* silencing in a murine AML model led to a decrease of immunophenotypically-defined LSCs, decrease in quiescent LSCs, and reduced clonogenic potential of LSCs [437]. *INPP4B* encodes for an enzyme involved in phosphatidylinositol dephosphorylation; it negatively regulates PI3K/AKT signalling and promotes lysosome biogenesis [434, 440]. High expression of *INPP4B* is associated with chemotherapy resistance, whilst overexpression in leukaemic cell lines enhanced colony forming potential [438]. Moreover, deletion of *INPP4B* in murine *MLL-AF9* leukaemic cells delayed leukaemia onset, decreased leukaemogenic potential in transplantation assays, and promoted differentiation *in vitro* [434]. This suggests that INPP4B may actively maintain an LSC state. Besides *SOCS2* and *INPP4B*, *NR4A2*, a gene bound by PRC1.1 in 4/6 of analysed human AML cases [367], was also more accessible in both *BCOR*-mutant clones. NR4A2 promotes HSC quiescence [36], suggesting TARB1/TARB2 LMPP-like cells may reside in a more quiescent state. Accordingly, chromatin more accessible in *BCOR*-mutant LMPP-like cells was enriched for genes highly expressed in LSCs/HSCs, compared to leukaemic/normal progenitors (Figure 5.25a). Consistent with *BCOR*-mutant LMPP-like cells being more quiescent, the promoter region and several distal elements co-accessible with *CDK20* (a positive cell cycle regulator) were less accessible in TARB2 LMPP-like cells (Figure 5.25b). Hence, epigenetic de-regulation of specific genes promoting stem-like states may mediate the retention of *BCOR*-mutant clones within LSCs.

Of note, the larger *BCOR*-mutant clone (TARB2) also displayed higher accessibility of *PBX3*, *JMJD1C*, *BTG3*, and *NPR3* (Figure 5.24b), putative PRC1.1 targets related to quiescence and/or promoting leukaemogenesis [231, 418, 419, 441–445].

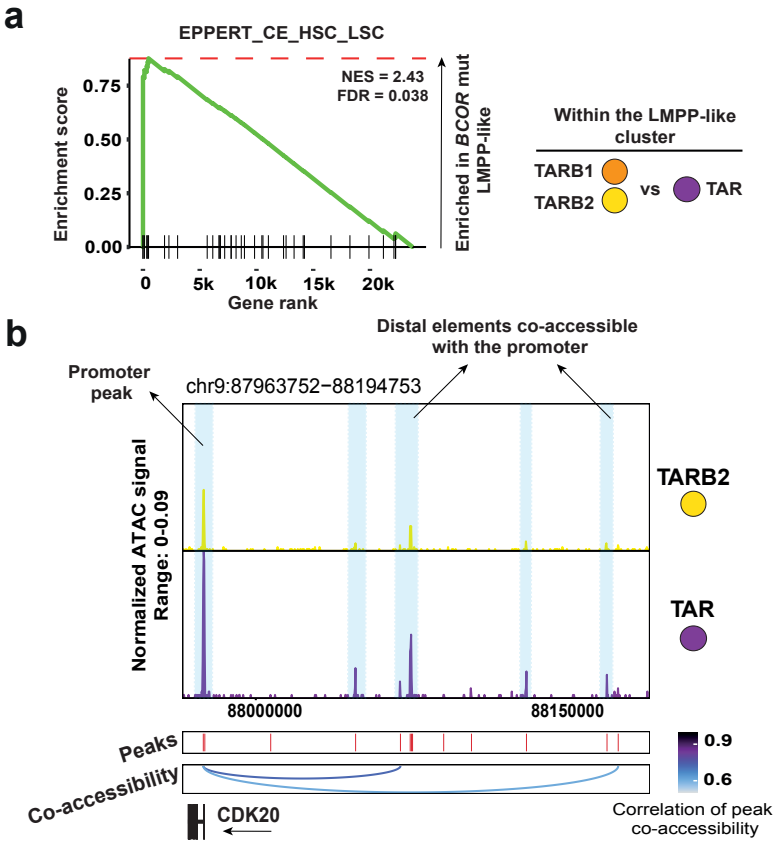


Figure 5.25: Chromatin characteristics of *BCOR*-mutant LMPP-like cells. **(a)** Enrichment of an HSC/LSC gene signature [231] in genes more accessible within *BCOR*-mutant versus *BCOR*-WT LMPP-like cells. **(b)** Chromatin accessibility around *CDK20*, comparing TARB2 and TAR LMPP-like cells.

In particular, PBX3 promotes LSC maintenance and cooperates with MEIS1, whilst JMJD1C has been functionally implicated in self-renewal of *MLL*r LSCs. Indeed, an interesting observation emerging from these analyses was that genes with increased accessibility in *BCOR*-mutant LMPP-like cells were reminiscent of a gene program downstream of *MLL*r. *MLL*r AMLs show a characteristic expression profile, with up-regulation of stemness-related genes, typically expressed within immature haematopoietic cells. Examples are *SOCS2*, *PBX3*, *JMJD1C*, *SKIDA1* [201, 208, 444, 446], all direct targets of *MLL*r. Moreover, *MLLT10* (AF10), also more accessible in TARB2 LMPP-like cells, supports aberrant gene transcription in *MLL*r AML [447]. These observations suggest that loss of PRC1.1. may mimic *MLL*r activity.

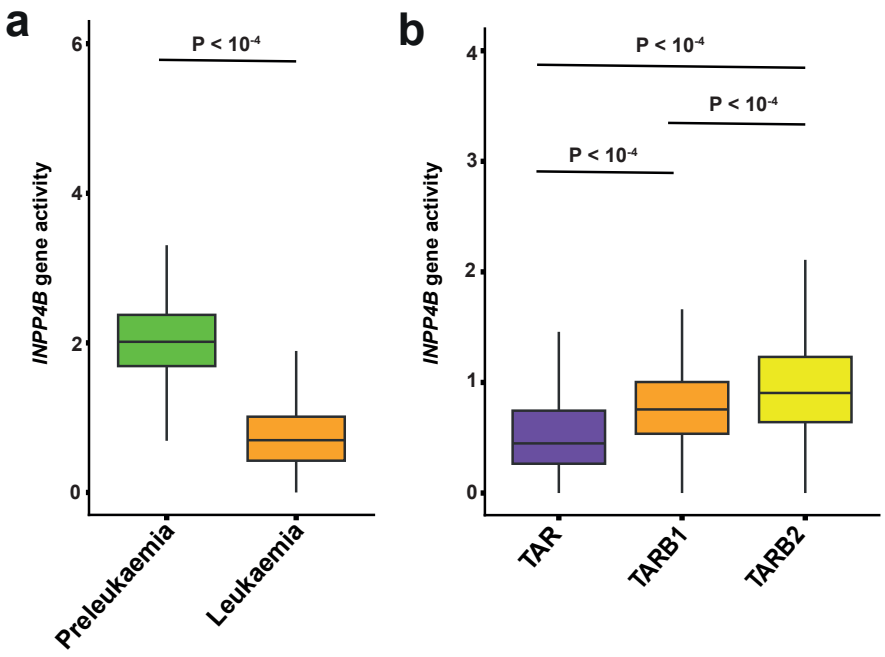


Figure 5.26: Clone-specific *INPP4B* accessibility. **(a)** *INPP4B* accessibility between pre-leukaemic (T/TA) and leukaemic (TAR/TARB1/TARB2) cells. Wilcoxon rank sum test. **(b)** *INPP4B* accessibility between TAR, TARB1 and TARB2 clones. Kruskal-Wallis test with Dunn's multiple comparisons.

I next asked whether some of the above findings may explain the frequent co-occurrence of *BCOR* and *RUNX1* mutations. Interestingly, two studies suggested that *INPP4B* is positively regulated by RUNX1 [88, 448]. Consistent with this, *INPP4B* was less accessible in *RUNX1*-mutant pre-leukaemic versus *RUNX1*-WT leukaemic cells (Figure 5.26a). Conversely, within the leukaemic hierarchy, *INPP4B* was more accessible in TARB clones (Figure 5.26b). This raises the hypothesis that PRC1.1 loss rescues *INPP4B* expression in *RUNX1*-deficient cells, which may promote maintenance of *BCOR*/*RUNX1*-mutant LSCs. This is in line with recent data from the Majeti group showing that *BCOR* deficiency rescues stemness programs in human *RUNX1*-deficient $CD34^+$ cells, where stem cell function is compromised [151, 319]. Interestingly, *INPP4B*, *NR4A2*, *MLLT10*, *BTG3*, and *SOCS2* were more accessible in *BCOR*-mutant clones as they differentiated throughout the leukaemic hierarchy (Figure 5.27), suggesting that inefficient repression of these genes within LMPP-like cells is propagated throughout differentiation of *BCOR*-mutant clones.

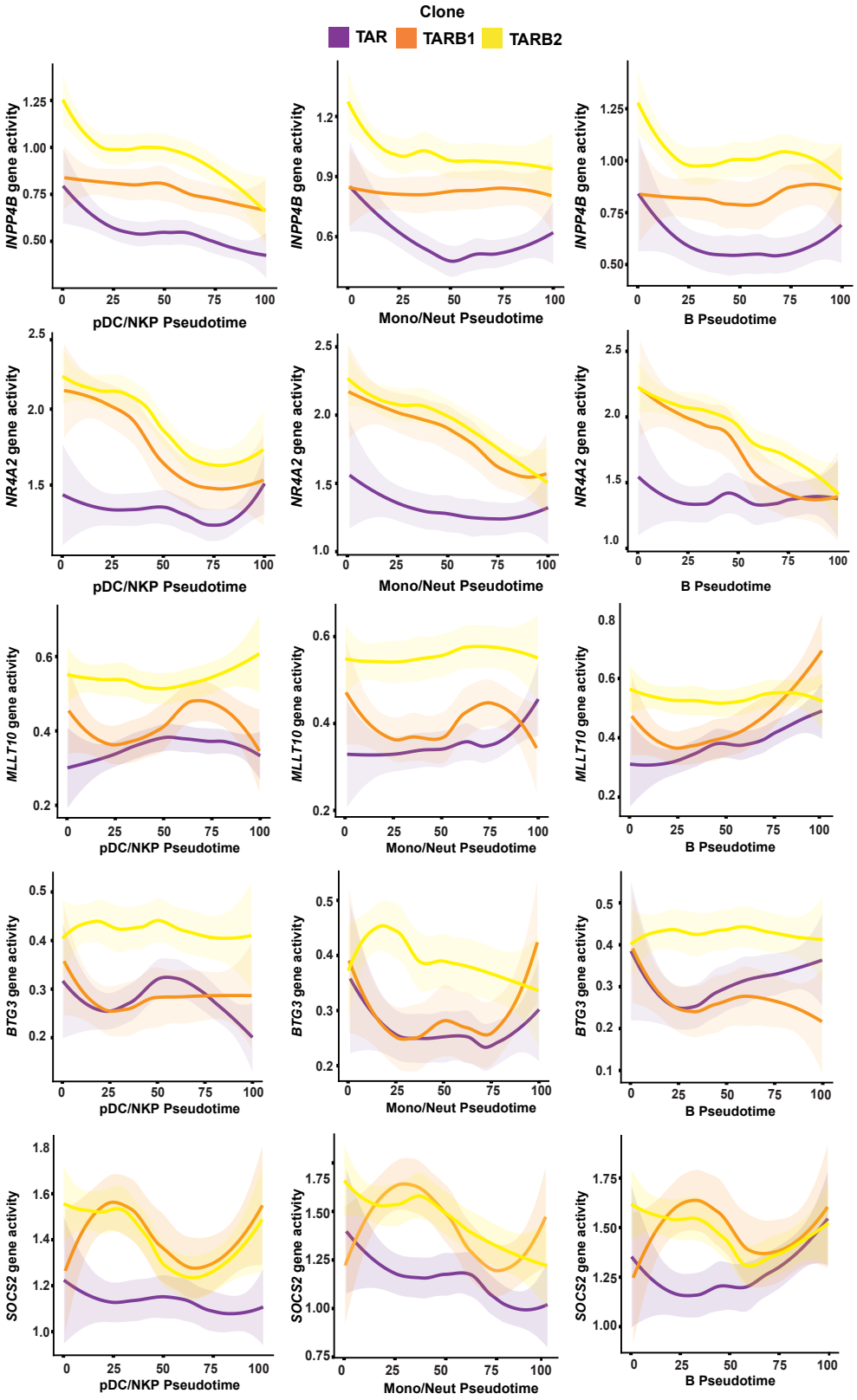


Figure 5.27: *Legend on next page*

Figure 5.27 *(previous page)*: Clone-specific gene accessibility through pseudotime. *INPP4B*, *NR4A2*, *MLLT10* (AF10), *BTG3*, and *SOCS2* gene activity (y-axis) across pseudotime (x-axis) within three leukaemic trajectories (columns). Clones are represented by colour-coded lines. The shadowed area indicates the CI of the loess method.

Altogether, *BCOR* LoF mutations in this somatic context correlate with epigenetic de-regulation of genes involved in self-renewal, quiescence, and LSC maintenance, within molecularly defined LSCs. Failure to repress genes like *INPP4B* may explain the increased fitness of *RUNX1*/*BCOR*-mutant clones.

5.5.2 *BCOR* mutations de-regulate promoters of LSC-associated genes

Next, I asked by which mechanism PRC1.1 loss de-regulated stemness-associated genes. Previously, it has been shown that PRC1.1 binding to chromatin has a substantial bias for CpG islands at TSS-proximal regions [367, 428], in contrast to PRC1, which binds predominantly to intronic or intergenic regions [367]. This suggests PRC1.1 regulates gene expression via deposition of H2AK119Ub at promoters.

Indeed, peaks more accessible in *BCOR*-mutant LMPP-like cells were enriched for promoter over distal elements (Figure 5.28a). This was significant for TARB2 but not TARB1 cells, likely due to the low numbers of promoter and distal peaks differentially accessible between TARB1 and TAR LMPP-like cells. These data are consistent with absent repression of specific promoter regions, given the failure to deposit H2AK119Ub1 at promoters in PRC1.1-deficient cells. Accordingly, the TSS-proximal region of *INPP4B* was more accessible in *BCOR*-mutant LMPP-like cells (Figure 5.28b-c). Differences in accessibility at TSS-proximal regions were also observed between TARB2 and TAR LMPP-like cells at several other de-regulated genes involved in quiescence and AML maintenance (Figures 5.29a-d). This held true both for genes more (*BTG3* and *SOCS2*) and less (*MLLT10* and *NR4A2*) frequently bound by PRC1.1 in human AML. Failure to repress these gene promoters may relate to retained accessibility of these genes as *BCOR*-mutant clones differentiate (Figure 5.27). Hence, GTAC suggested a mechanistic link between *BCOR* mutations and epigenetic de-regulation.

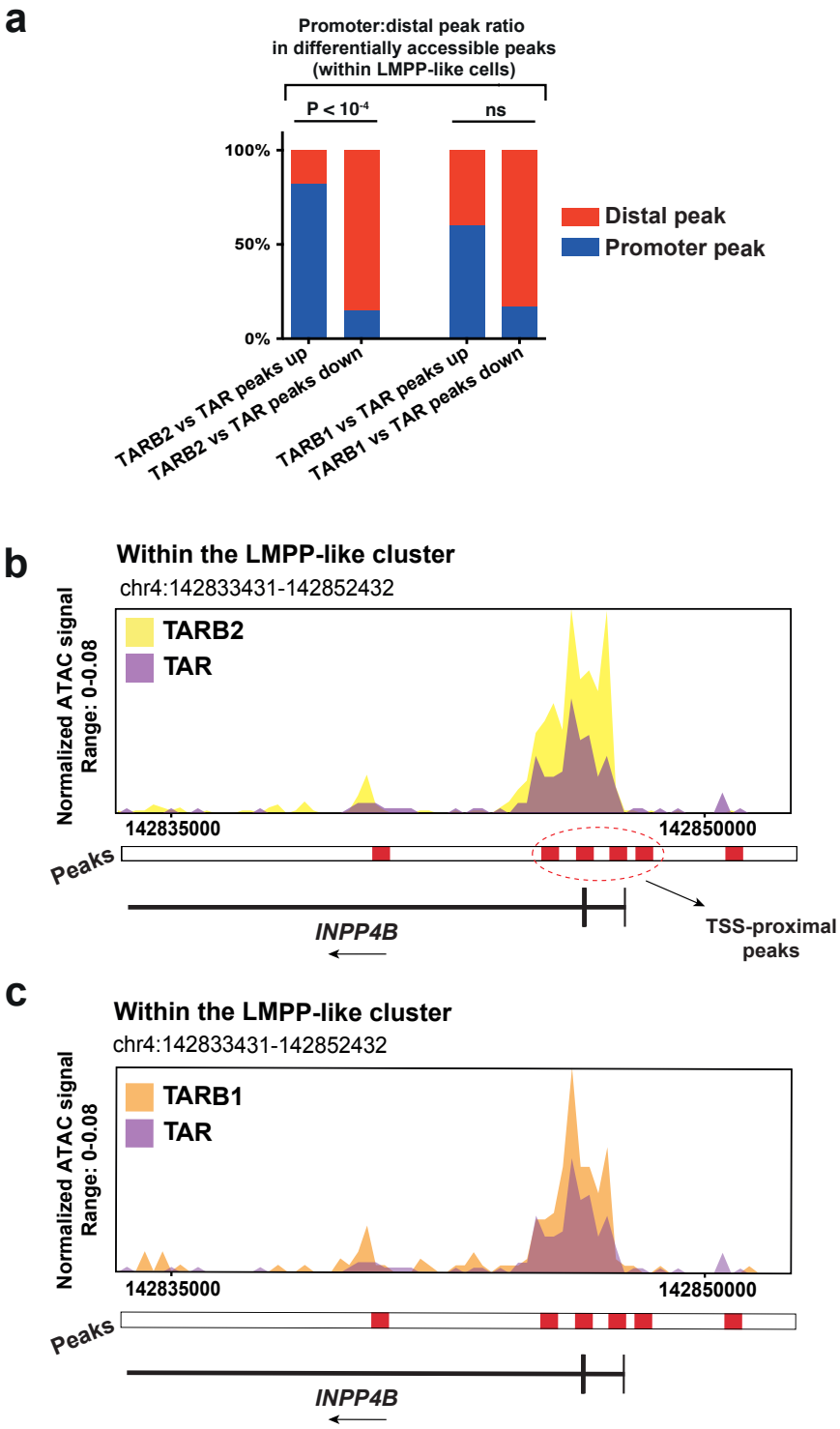


Figure 5.28: *INPP4B* promoter de-regulation in *BCOR*-mutant LMPP-like cells. (a) Ratio of promoter versus distal elements peaks, within peaks which are either significantly more or less accessible in *BCOR*-mutant versus *BCOR*-WT LMPP-like cells (FDR < 0.05 and absolute $\log_2$FC > 0.25). Fisher's exact test. **(b)** Chromatin accessibility around the *INPP4B* promoter in TAR (purple) and TARB2 (yellow) LMPP-like cells. **(c)** As in (b) but between TAR and TARB1.

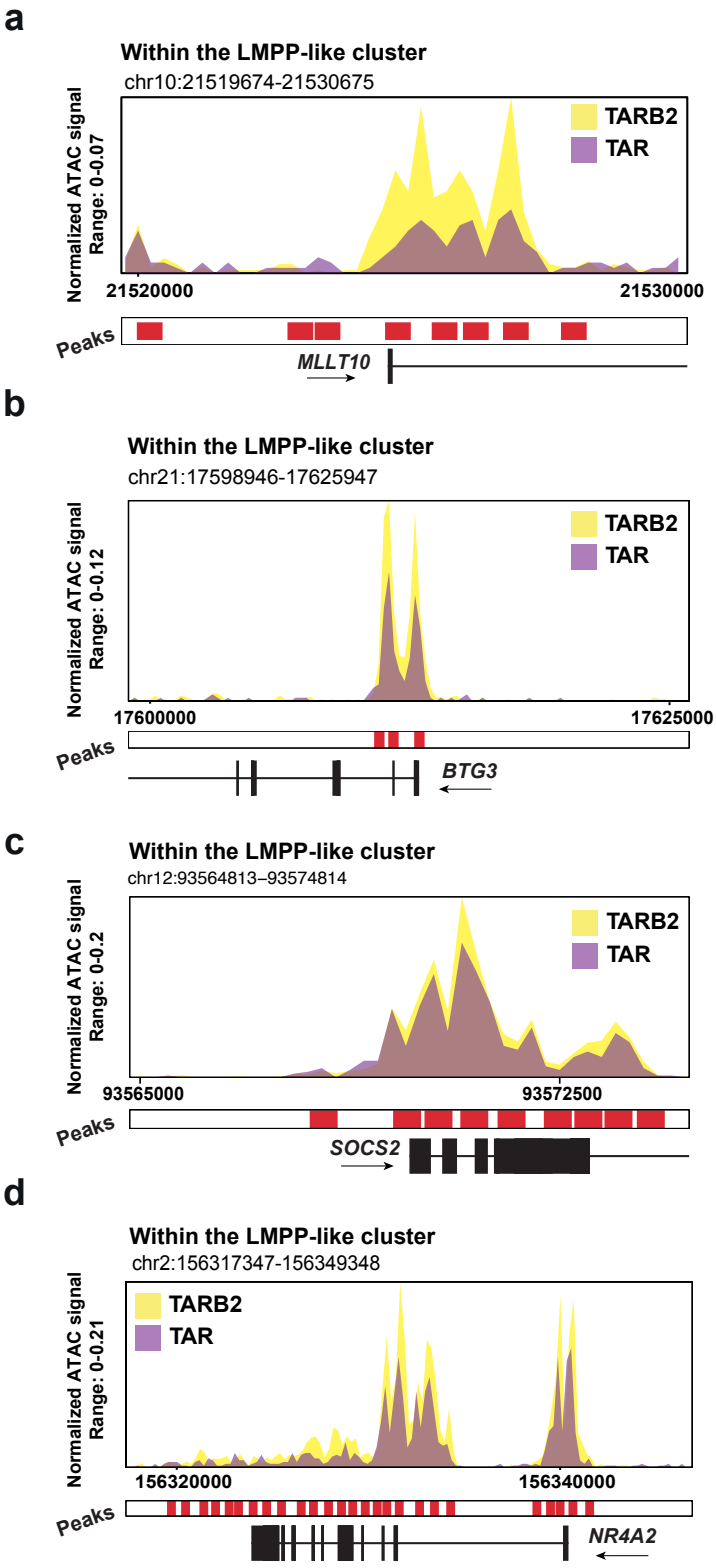


Figure 5.29: Promoter de-regulation in *BCOR*-mutant LMPP-like cells. **(a)** to **(d)** As in Figure 5.28b but around the *MLLT10*, *BTG3*, *SOCS2*, and *NR4A2* promoters.

5.5.3 Aberrant TF activity in *BCOR*-mutant LMPP-like cells

I next investigated putative TFs driving increased retention of stemness programs in concert with PRC1.1 loss. Comparison of TF binding motifs between *BCOR*-mutant and *BCOR*-WT LMPP-like cells (Figures 5.30a-b) revealed shared patterns but also surprising differences between the two *BCOR*-mutant clones.

Both TARB1 and TARB2 LMPP-like cells had higher accessibility of motifs bound by HOX and SMAD families (Figures 5.30a-b), associated with stemness, quiescence, and CSC function. Accordingly, experimental *BCOR* deletion in haematopoietic cells correlates with increased *HOX* gene expression [424, 425]. HOX family TFs are developmental regulators frequently overexpressed in AML [449]. In normal haematopoiesis, selected *HOXA*/*HOXB* genes are highly expressed in primitive HSPCs and downregulated during differentiation [449]. In AML, high expression of some *HOX* genes in progenitor-like cells associates with aberrant self-renewal programs which support LSC phenotypes. Accordingly, high *HOXA* cluster expression is related to poor prognosis in AML [450, 451]. Binding of HOX TFs to DNA is dependent on MEIS1 and PBX family co-factors; consistently, *PBX3*, encoding a co-factor of HOXA9 [418], was more accessible in TARB2 LMPP-like cells (Figure 5.24b). *HOX* genes are positively regulated by MLL [53, 55] and are highly expressed in *MLL*r AML [201, 446, 452]. Hence, high HOX binding motif accessibility in *BCOR*-mutant cells is in line with their resemblance to *MLL*r LSCs.

Compared to TAR, TARB1 LMPP-like cells showed higher accessibility of motifs bound by ETV6, ERG, and FLI-1 (Figure 5.30a), ETS family of TFs implicated in maintenance of HSCs and LSCs. ETV6 is essential for the survival of murine HSCs, but not downstream progenitors [24]. Similarly, ERG promotes HSC quiescence [25, 417, 453] via activation of self-renewal genes, whilst inhibiting HSC differentiation via repression of MYC targets. FLI-1, a binding partner of ERG, is also indispensable for normal HSC maintenance [26]. Motifs bound by FOXO TFs, which maintain HSC quiescence [44], were more accessible in TARB1 LMPP-like cells (Figure 5.30a), in line with FOXO1 upregulation in *BCOR*-mutant versus *BCOR*-WT AML patients [432]. Finally, TARB1 LMPP-like cells had increased accessibility of motifs bound

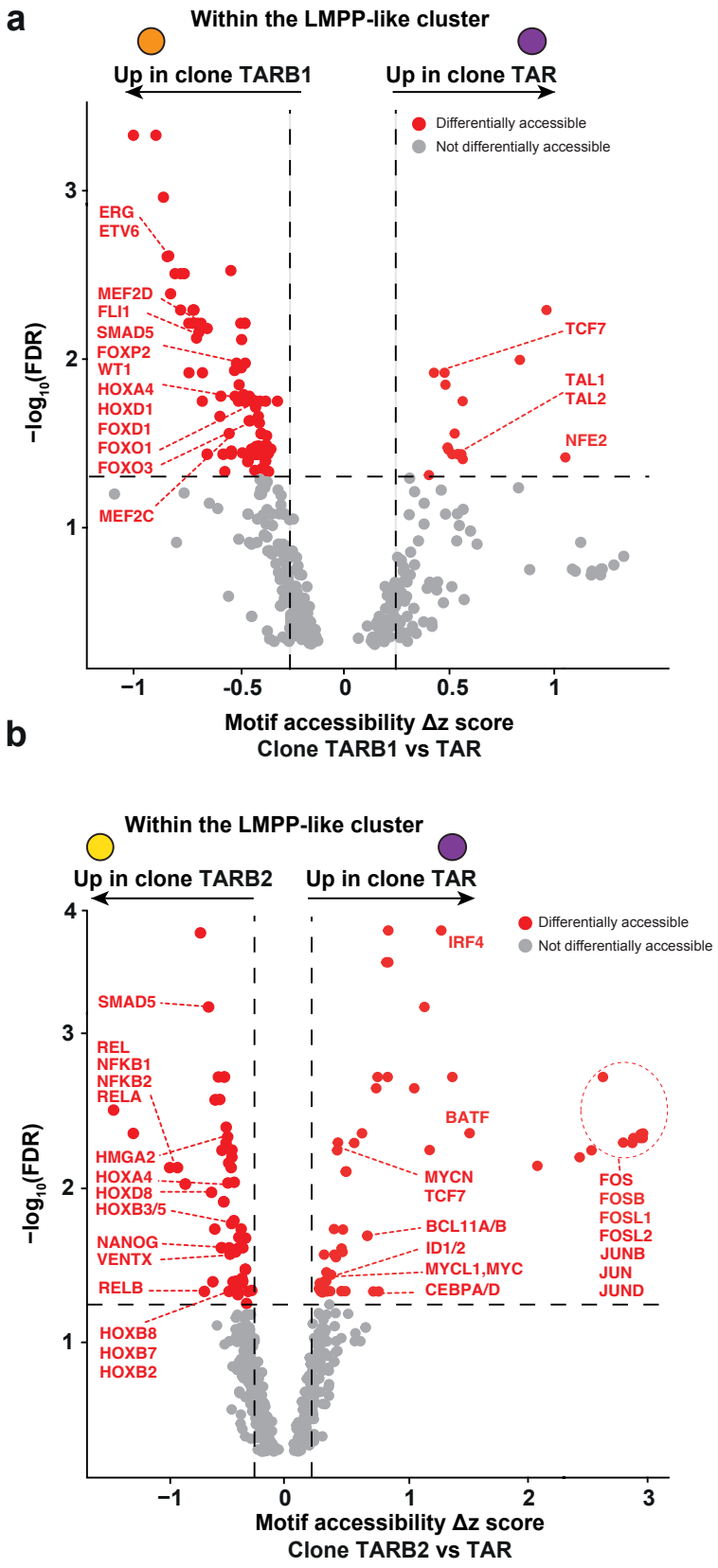


Figure 5.30: *Legend on next page*

Figure 5.30 *(previous page)*: TF binding motif accessibility between *BCOR*-mutant and *BCOR*-WT LMPP-like cells. (a) Differentially accessible TF binding motifs between TARB1 (n=130 cells) and TAR (n=167 cells) LMPP-like cells (FDR < 0.05 and absolute $\Delta z > 0.25$; dashed lines). **(b)** As in (a) but comparing TAR (n=167 cells) and TARB2 (n=293 cells) LMPP-like cells.

by MEF2C and MEF2D (Figure 5.30a). *MEF2C* is another target of *MLL*r [208, 446], important for AML chemoresistance [403], and highly expressed in LSCs [244]. Interestingly, MEF2C positively regulates *SOCS2* transcription [435], consistent with higher *SOCS2* accessibility in *BCOR*-mutant LMPP-like cells (Figure 5.24). Both *MEF2C* and *SOCS2* are part of a *MECOM*-regulated stemness gene program [30], in line with increased stemness of *BCOR*-mutant LMPP-like cells.

Surprisingly, although TARB2 LMPP-like cells also showed enhanced accessibility of HOX motifs, there were differences compared to the TARB1 clone. Specifically, compared to TAR, TARB2 LMPP-like cells were enriched for motifs bound by NFKB family, HMGA2, and the pluripotency-associated TF NANOG (Figure 5.30b). HMGA2 is typically expressed in most immature AML cells [454] and maintains HSC function [455]. NF kB, TFs typically associated with inflammatory signalling, are highly expressed in primitive $CD34^{+}CD38^{-}$ LSCs and are required for their survival [250]. It is also notable that TNFα signalling via NF kB is enriched in dormant leukaemic cell fractions [258] and in a subset of HSCs [79, 80].

Conversely, TAR LMPP-like cells were enriched for motifs bound by TFs implicated in differentiation, like TCF7, NFE2, IRF4, BCL11A/B, and CEBPA/D (Figures 5.30a-b). Moreover, motifs bound by the AP-1 complex and by MYC, MYCN, and MYCL were more accessible in TAR LMPP-like cells (Figure 5.30b). These observations further suggest that *BCOR*-WT LMPP-like cells are epigenetically primed to cycle and differentiate, while their *BCOR*-mutant counterparts are increasingly primed towards self-renewal (Figure 5.25a).

Of note, differences between *BCOR*-mutant and *BCOR*-WT leukaemic cells were most pronounced within the LMPP-like cluster. Similar comparisons between TARB1/2 and TAR clones within more differentiated Mono-Neut progenitor-like

cells yielded less differences (Figures 5.31a-b). CEBP family, IRF4, and BCL11A/B motifs were still less accessible in *BCOR*-mutant cells (Figure 5.31a), supporting imapired myeloid differentiation of TARB2 cells (Figures 5.20, 5.21, and 5.23).

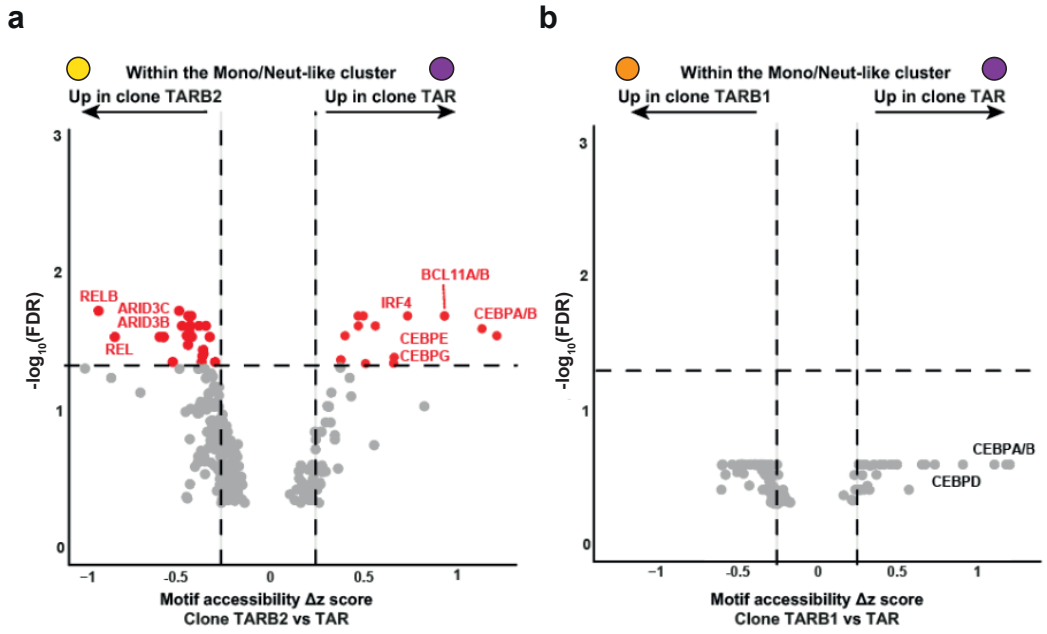


Figure 5.31: TF binding motif accessibility between *BCOR*-mutant and *BCOR*-WT Mono/Neut progenitor-like cells. (a) Differentially accessible TF binding motifs between TARB2 (n=175 cells) and TAR (n=250 cells) Mono/Neut progenitor-like cells (FDR < 0.05 and absolute $\Delta z > 0.25$; dashed lines). **(b)** As in (a) but comparing TARB1 (n=128 cells) and TAR (n=250 cells) Mono/Neut progenitor-like cells.

There were no significant differences between TARB1 and TAR Mono/Neut progenitor-like cells (Figure 5.31b). It is thus likely that the main consequences of *BCOR* mutations occurred within LMPP-like cells.

Finally, I wanted to validate the robustness of the above analyses. Given that a small proportion of cells labelled as TAR had failed genotyping at the *BCOR* locus (Figure 5.7a), it is to be expected that a small proportion of 'TAR' cells are, in fact, TARB1/2 cells, in which the mutant allele was not sampled. To this end, I purposefully misassigned 5% of TARB2 LMPP-like cells as TAR and repeated the analyses from Figures 5.24b, 5.30b, and 5.28b.

Encouragingly, results were comparable (Figures 5.32a-c), albeit with reduced level of significance for some genes and TF motifs. This indicates that modest levels of clone mis-assignment lead to a subtle dilution of the true differences between clones.

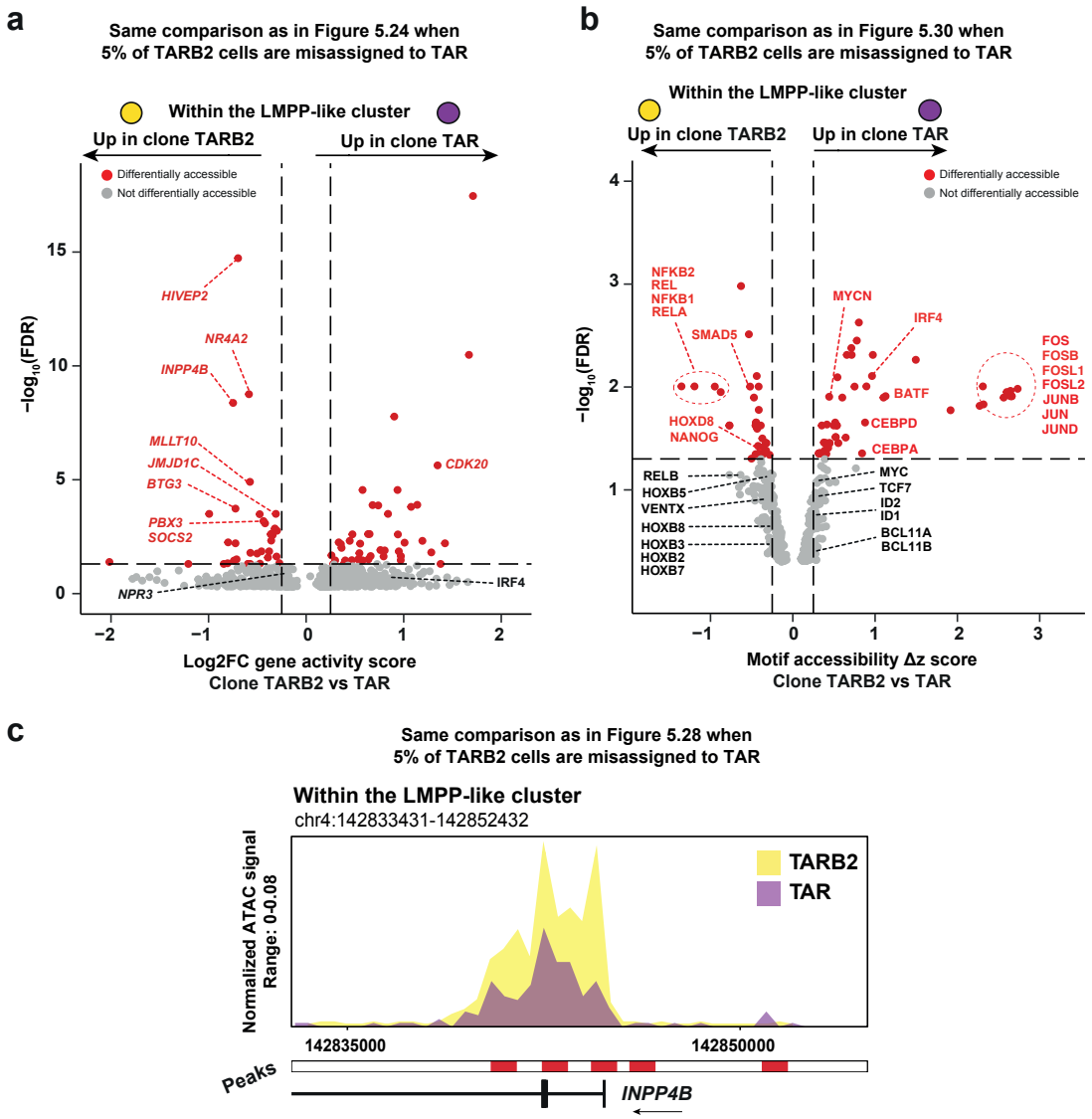


Figure 5.32: Simulation of ADO effect on differential accessibility analysis. **(a)-(c)** Same plots as in Figures 5.24b, 5.30b, and 5.28b, when 5% of random TARB2 cells were purposefully misassigned as TAR.

Taken together, the data above suggest failure to repress stemness-related TF programs in *RUNX1/BCOR*-mutant LSCs. Higher epigenetic priming of *BCOR*-mutant cells towards self-renewal may account for their clonal selection.

5.5.4 Epigenetic heterogeneity between genetically similar clones

Whilst uncovering important shared patterns, the above analyses alluded to an unexpected degree of epigenetic heterogeneity between the two *BCOR*-mutant subclones. This was somewhat surprising, as both *BCOR* mutations are predicted

LoF. Hence, this dataset provided the unique opportunity to study epigenetic differences between two clones with apparently similar genotypes, within the same patient.

Comparison of all TARB1 and TARB2 leukaemic cells revealed several differentially accessible genes (Figure 5.33). Genes which were more accessible in the larger TARB2 clone are linked to stemness, self-renewal, and AML cell maintenance, like *HOXA9*, *HOXA10*, *KIT*, *BMI1*, *NR4A2*, and *MLLT10*. *HOXA9* and *HOXA10* encode for key AML oncogenes [456–460]. Interestingly, *INPP4B* was also more accessible within TARB2 cells. Differential TF binding motif accessibility (Figure 5.34) showed increased HOX motif openness in TARB2 cells, whereas motifs for differentiation-associated BCL11A/B, IRF4, and SPI1 TFs were more open in TARB1 cells. Hence, whilst both *BCOR*-mutant clones share epigenetic features of primitive stem-like states, they were not epigenetically equivalent. In fact, the TARB2 clone likely occupies a more primitive stem-like state within the leukaemic hierarchy.

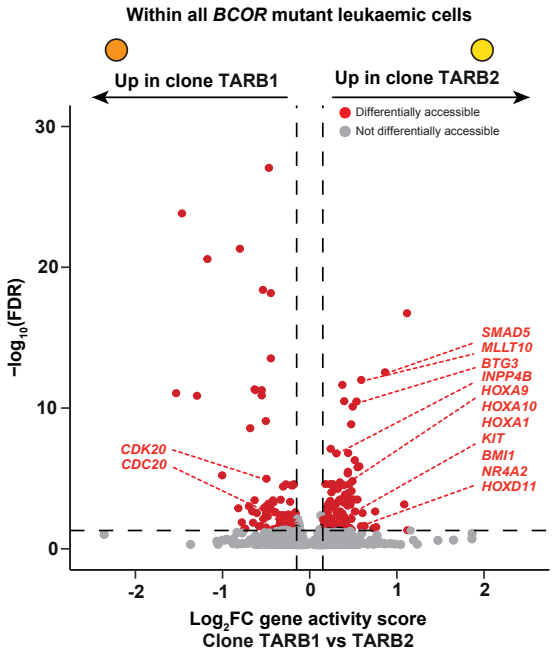


Figure 5.33: Differential gene accessibility between *BCOR*-mutant clones. Differentially accessible genes between all TARB1 (n=446 cells) and TARB2 (n=834 cells) leukaemic cells (FDR < 0.05 and absolute $\log_2$FC > 0.20; dashed lines).

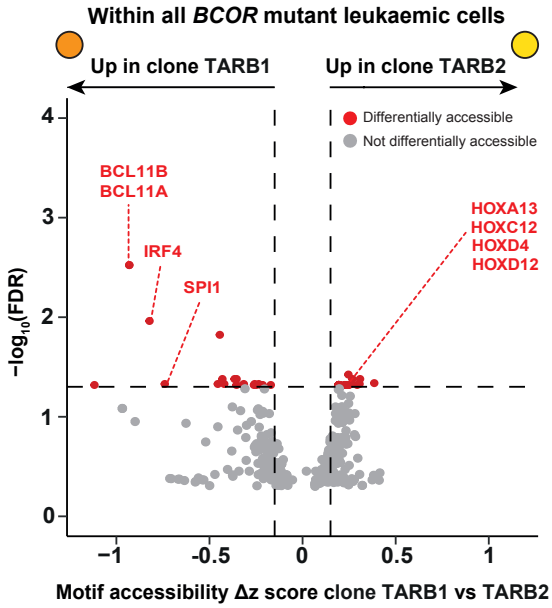


Figure 5.34: Differential TF binding motif accessibility between the two *BCOR*-mutant clones. Differential accessibility of TF binding motif between TARB1 (left) and TARB2 (right) leukaemic cells. Differentially accessible motifs, defined as those displaying FDR < 0.05 and absolute $\Delta z > 0.20$, are labelled as red dots.

This is reflected in distinct differentiation dynamics between the two clones (Figure 5.20), where the TARB2 clone was more delayed in differentiation compared to TARB1. This, in turn, may explain its larger size. These data demonstrate that genetically similar clones can be epigenetically heterogeneous, an emerging concept potentially generalizable across the broader field of cancer biology.

5.6 Brief summary of Chapter 5

GTAC enabled precise mapping of clonal progression to epigenomic evolution in human AML (Figure 5.35). In this patient, a homozygous *RUNX1* LoF mutation was the key genetic event, transforming genetically pre-leukaemic cells into malignant leukaemic cells. Accordingly, *RUNX1*-associated leukaemic transformation, which occurred at the LMPP stage, was associated with global pro-proliferative epigenomic changes. GTAC further revealed more subtle switches in gene regulation as leukaemic cells clonally evolved by acquiring LoF *BCOR* mutations. The resulting *BCOR*-mutant leukaemic progenitors differentiated less efficiently, instead accumulating within a differentiation stage resembling LSCs. I show that *BCOR*-mutant LSCs

display absent repression of stemness-associated programs due to the absence of PRC1.1 activity at promoters of key genes. Hence, *BCOR*-mutant cells, primed towards self-renewal and against differentiation, gradually expanded the LSC pool within this patient. By comparing *BCOR*-mutant and *BCOR*-WT molecular LSCs, I identified potential therapeutic targets, the exemplar being *INPP4B*, which warrant further investigation. Importantly, regulation of *INPP4B* may, at least in part, explain the cooperativity between *RUNX1* and *BCOR* mutations. Finally, the unexpected epigenetic heterogeneity arising between the two *BCOR*-mutant clones suggests that even apparently similar genotypes can show a degree of epigenetic heterogeneity. Whilst these experiments demonstrated that GTAC can link somatic mutations to epigenetic variation, definitive conclusions warrant validation in a larger number of patients, followed by functional testing of the identified targets. Eventually, these studies have the potential to identify more efficient therapeutic options for this patient subgroup. More broadly, this study served as a proof-of-principle that GTAC can be adopted to gain mechanistic insights into how somatic mutations generate and maintain LSC phenotypes, leading to potential therapeutic insights.

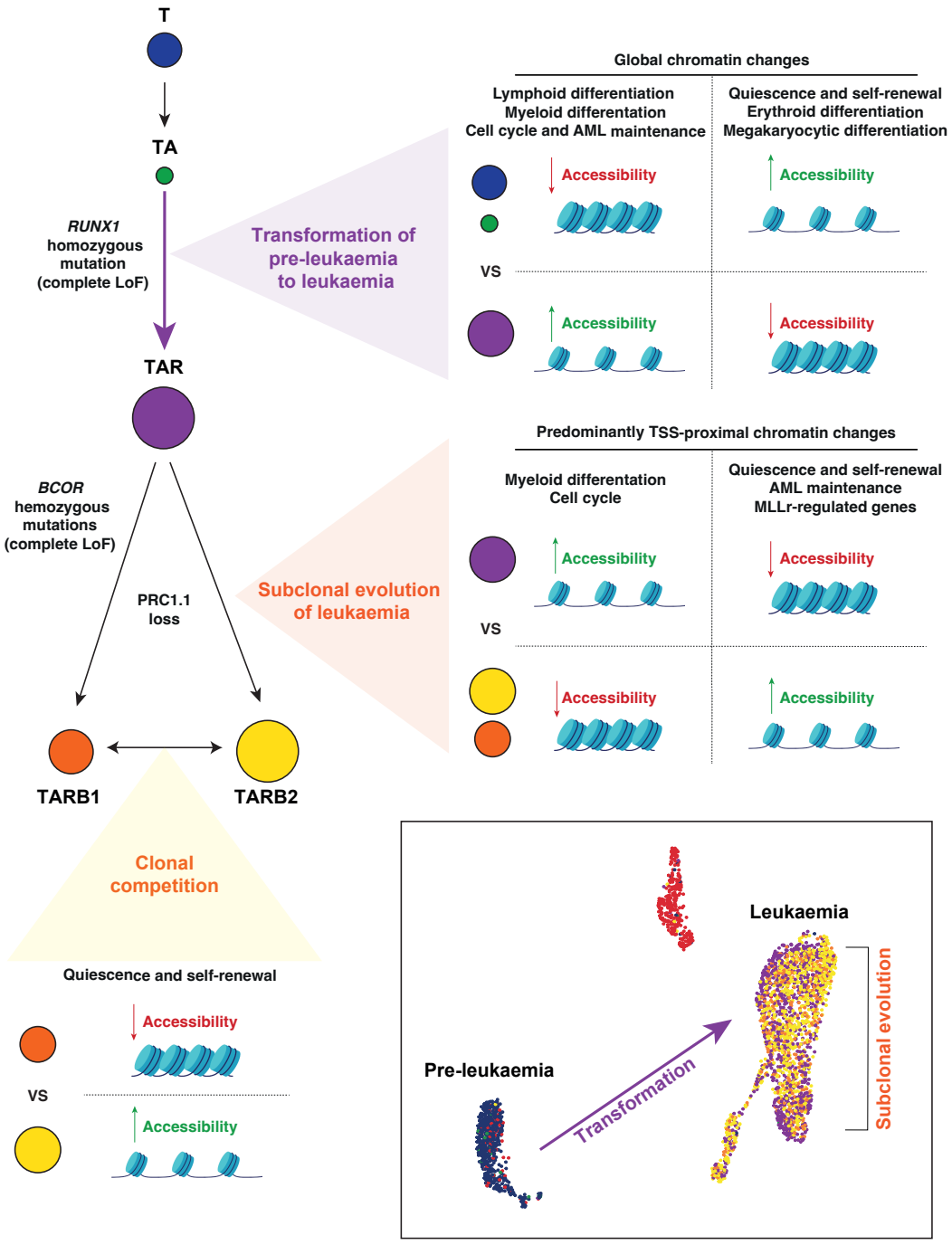


Figure 5.35: Clonal evolution correlates with epigenetic change. The key findings emerging from GTAC analysis.

Chapter 6

Discussion

Contents

6.1 The implications of clonal heterogeneity for AML treatment 182
6.2 Clonal and cellular basis of resistance and response to ivosidenib combination therapies 185
6.2.1 Rapid clonal selection within LSCs foreshadows patient outcome 185
6.2.2 Heterogeneous genetic and cellular routes towards relapse 187
6.2.3 Variable duration of clinical response despite universally rapid clonal selection 188
6.2.4 Implications for clinical practice 189
6.2.5 Challenges ahead and future outlook 190
6.3 GTAC: integrating clonal identity with chromatin accessibility 191
6.3.1 Technical improvement of GTAC over prior approaches 191
6.3.2 Insights into *RUNX1*-associated leukaemic transformation 192
6.3.3 The impact of somatic mutations on LSC gene regulation 194
6.3.4 Future applications of GTAC 196
6.3.5 Technical considerations, limitations, and future outlook 198

6.1 The implications of clonal heterogeneity for AML treatment

Despite decades of research, survival of AML patients, particularly those ineligible for intensive chemotherapy, has improved only modestly from the 1970s to the present day [167]. Hence, devising efficient AML therapies remains a major unmet clinical need. Dismal outcome is primarily linked to acquired treatment resistance and relapse. This warrants a more precise understanding of the fundamental biology of human AML, with the scope of devising more efficient treatment regimens. Specifically, how does resistant disease bypass treatment and what are the genetic and molecular determinants of treatment failure, remain open questions. Whilst

treatments can initially suppress malignant AML cell populations, they concomitantly provide a unique selective pressure for the Darwinian outgrowth of resistant cells. Accordingly, relapse frequently correlates with the selection of genetic clones harbouring resistance-associated mutations [174], or, alternatively, with putative epigenetic adaptation to treatment [314]. Clonal selection presumably occurs within cells harbouring LSC potential, capable of re-establishing malignant outgrowth despite treatment. Hence, to tackle the issue of relapse, a deep understanding of both genetic and non-genetic drivers of LSC phenotypes is required.

Key experimental limitations have prevented a complete picture of treatment resistance in AML. Apart from rare cases in which a single genetic driver is present, AML is frequently composed from multiple heterogeneous clones, in which cooperating lesions disrupt normal blood production. These clones exert differential potential to resist therapy and change dynamically throughout treatment. At present, such complexity is hard, if not impossible, to model in experimental systems, and this likely applies to other cancers beyond AML. Whilst providing important insights about the consequences of mutations, functional approaches have so far not fully elucidated the combinatorial effects downstream of multiple somatic events in a complex human clonal milieu. Moreover, malignant clones are likely to arise from a years-long selective process, mediated by heterogeneous environmental stimuli. For example, clonal advantage of *DNMT3A/TET2* pre-leukaemic cells or *TP53*-mutant leukaemic cells is driven by external inflammatory pressures [80, 289–292, 331]. In turn, these inflammatory stimuli can be related to infections from various sources, autoimmune diseases, therapeutic interventions, or even to a specific diet. In the case of lung adenocarcinoma, it was proposed that air pollutants promote the growth of malignant mutant clones via macrophage-mediated inflammation [461]. I speculate that, in most cancers, clonal advantage of pre-malignant and malignant cells likely results from a complex interaction between germline genetics, somatic genetic change, and environmental effects. In this regard, inbred model organisms which are maintained in controlled environments may be limited in the

ability to accurately reproduce the breadth of clonal and environmental complexity promoting human cancer.

For these reasons, the aim of this thesis was to improve our understanding of the clonal and molecular landscapes in primary human AML. Three aspects were addressed:

1. The rapidity and the cellular context of acquired resistance to a specific AML treatment were investigated, to provide fundamental insights about treatment failure, likely generalizable across most AML therapies.

2. Methods to study how mutations cooperatively alter gene regulation, leading to clonal advantage in primary human disease samples, are required.

3. A precise characterization of genetic and epigenetic factors promoting LSC maintenance and expansion are warranted to devise LSC-targeting therapies.

Briefly, the key novel findings presented in this thesis are:

1. Rapid selection of genetic clones resistant to ivosidenib combination therapies occurs specifically within small BM populations with LSC potential. Selection is operative while patients are clinically responding to therapy, months or years prior to clinically overt relapse.

2. A method for simultaneous high-confidence single-cell genotyping of genomic loci and chromatin accessibility profiling in primary human samples was developed.

3. A *RUNX1* LoF mutation associates with leukaemic transformation of pre-leukaemia.

4. *BCOR* mutations associate with epigenetic dysregulation of specific stemness-associated genes and TF networks, which support LSC phenotypes. These mechanisms likely underly *BCOR*/*RUNX1* mutation cooperativity.

Below, I discuss the main implications and limitations of my results. Although distinct, the two projects presented here share the common goal of characterizing the clonal and molecular heterogeneity underpinning human AML, with important therapeutic implications.

6.2 Clonal and cellular basis of resistance and response to ivosidenib combination therapies

6.2.1 Rapid clonal selection within LSCs foreshadows patient outcome

By studying longitudinal samples from *IDH1*-mutated patients treated with ivosidenib and venetoclax ± azacitidine, I uncovered that universally rapid clonal selection correlated with longer-term patient outcome. Selection of leukaemic clones associated with relapse, or WT and clonal haematopoiesis sustaining durable remission, occurs within 3 cycles of treatment.

In patients who relapsed, clonal selection preceded treatment failure by months or even years. Hence, resistant clones are detectable while treatment still efficiently suppresses malignant proliferation. Relapse was associated with somatic lesions previously related to ivosidenib and venetoclax resistance. Emergence/selection of receptor tyrosine kinase (RTK) pathway mutations (like *FLT3* or *NF1*), or TF mutations (*RUNX1*), correlate with relapse following inhibition of mutant IDH [270, 302, 462]. Similarly, RTK pathway mutations confer venetoclax resistance [305, 463]. RAS/MAPK, PI3K/AKT, and STAT5 pathways, downstream of FLT3, upregulate BCL-xL and MCL-1, which compensate for BCL-2 inhibition [464] (reviewed in [307]. Conversely, the role of CNAs selected at relapse is less clear, likely owing to the difficulty of modelling these aberrations. In general, complex karyotype in AML associates with dismal prognosis [172]. Of note, deletions of chromosome 7 (pt4) associate with poor response to venetoclax [376], whilst trisomy 4 (pt14) associates with hypermethylation [373], potentially conferring ivosidenib resistance. Whilst the above somatic lesions were previously associated with resistance to this combination therapy [199], my results provide novel insights into the kinetics, timing, and cellular context of selection. The dominant relapse clone was always

distinct from the dominant diagnostic clone, regardless of whether newly detected mutations emerged or not at relapse. Two modalities of selection were observed: (a) resistant genetic clones, undetected at diagnosis, were selected upon treatment exposure; or (b) pre-existing clones, which were minor at diagnosis, rapidly lost differentiation capacity under therapeutic pressure, expanding at relapse. This indicates that dynamic clonal adaptation to treatment, which may or may not be genetically determined, presages treatment failure. Of note, 'newly emerging' clones either genuinely acquired an additional somatic mutation *de novo*, or were present already at diagnosis, but were undetectable by conventional approaches. The latter is more likely, as neither venetoclax nor ivosidenib are known to increase mutagenesis rates. In support of this are prior observations that relapse-associated clones often exist at very low frequency already at baseline [313, 465].

Crucially, regardless of the mechanism, rapid selection of resistant clones always occurred within small immunophenotypic BM populations with LSC potential. Selection was operative within LSCs whilst patients were clinically responding to treatment and myeloid differentiation was sustained by WT or genetically pre-leukaemic cells. This highlights an important phenomenon, where early selection within LSCs is a harbinger of relapse, despite the small size of the resistant clone at remission. Indeed, whilst resistant clones were often relatively small at EOC1/3 if Lin^- BM cells were analysed as a whole, at the same time these clones were present within minuscule LSC fractions at high frequency, sometimes exceeding 90%. Thus, uncontrolled cellular proliferation at relapse is preceded by a period of clonal accumulation within LSCs. Importantly, these findings suggest a potential temporal window for therapeutic intervention, aiming to eradicate mutant LSCs prior to the onset of clinical relapse (for example, via early administration of FLT3 inhibitors, in case *FLT3*-mutant clones start accumulating within LSCs). These findings build up on prior observations that failed eradication of immunophenotypic LSCs associates with relapse [257, 259], showing dynamic clonal re-organization within persistent LSCs. In agreement with the proposed model that early selection of resistant clones establishes the foundations of relapse, durable remission was

marked by early selection of clones harbouring only a known pre-leukaemic, but not transforming, genetic driver. Rapid selection of pre-leukaemic cells occurred in a background of WT haematopoiesis, whilst no resistant leukaemic clones were detected.

Thus, it is the clonal milieu at early remission which identified those patients destined to relapse.

6.2.2 Heterogeneous genetic and cellular routes towards relapse

Whilst rapid clonal selection was universally observed, my results highlight the heterogeneity of potential genetic and non-genetic routes leading towards ivosidenib + venetoclax ± azacitidine resistance.

Across patients, relapse-associated clones harboured distinct combinations of somatic mutations and were arrested in differentiation within differing BM populations. It is likely that specific combinations of mutations exhibit variable consequences within different stages of cellular maturation. Hence, further studies of resistant disease will have to consider both the clonal milieu and the differentiation stage at which resistant clones are arrested. The fact that both are highly heterogeneous between patients suggests acquired genetic resistance to this treatment is a complex clinical challenge, reminescent of genetic resistance to enasidenib [270], which requires increasingly personalized therapeutic approaches. Moreover, it is at present difficult to envisage how such inter-patient complexity can be faithfully reproduced in experimental settings. In this regard, future clone-aware comparisons of disease at relapse and diagnosis, carried out on a larger cohort, may reveal common altered pathways within different resistant clones across patients. Eventually, this may inform functional investigation of potential therapeutic targets. Interestingly, such genetic complexity is not necessarily observed across all AML treatments. For example, in *MLL*r or *NPM1*-mutated patients, genetic resistance to menin inhibition appears to be largely driven by mutations in a single gene, *MEN1* [308]. This suggests different therapies restrict malignant cells to a differing number of genetic bypass avenues. Moreover, these observations raise hopes that the pool

of potential resistance mechanisms hijacked by AML cells may be exhausted, by dynamically combining therapeutic agents in future trials.

An additional layer of complexity relates to putative non-genetic mechanisms of resistance. This is best exemplified by pt14 (Figure 3.12), where the pre-existing DN(+4) clone rapidly lost differentiation capacity upon treatment exposure, eventually expanding at relapse. One hypothesis is that therapy induced epigenetic reprogramming of DN(+4), causing this clone to differentiate less efficiently and expand within immature GM precursors. In fact, recent analyses suggest that ≈40% of AML patients relapse without apparent emerging driver mutations, accompanied by evolution at chromatin level [314]. Future linkage of somatic mutations and chromatin accessibility [332, 388] in longitudinal samples from these patients may thus be beneficial to uncover patterns of chromatin variation within resistant clones. Another, non mutually exclusive hypothesis, is that acquisition of unknown/undetected coding or non-coding genetic drivers may have occurred within these resistant clones. For example, genetic changes within the non-coding genome (occurring in cis-regulatory elements) may influence the activity of key TFs, thus disrupting differentiation of resistant clones. However, the non-coding genome is not routinely screened for mutation discovery. In this regard, longitudinal whole-genome sequencing would be helpful, although the distinction between positively selected non-coding drivers and non-driver passenger mutations is still not trivial [466].

The full spectrum of routes to resistance is likely to be greater than documented in these 6 relapsing patients and should be explored in future larger cohorts. These studies will also require adequate statistical power in order to identify potential systemic differences in resistance mechanisms between the ivosidenib/venetoclax doublet and the ivosidenib/venetoclax/azacitidine triplet regimen.

6.2.3 Variable duration of clinical response despite universally rapid clonal selection

Remarkably, despite early clonal selection, the DOR varied substantially between patients, ranging from months to years (Figure 3.1). This raises an obvious question: if clonal selection was universally rapid, why did some patients relapse only after

years on treatment? Several non-mutually exclusive hypotheses may explain this phenomenon. Firstly, different clones may exhibit variable fitness [467, 468] in this therapeutic setting, and some clones may have a faster doubling time than others. Secondly, differing BM microenvironments between patients may influence the differential kinetics of clonal outgrowth. Thirdly, clonal competition with alternative clones within a patient's BM may play a role in the relative fitness of a given resistant clone. In support of the latter, in pt5 and pt20, who stayed on treatment for > 2 years, resistant clones emerged rapidly, but competed with independent pre-leukaemic clones during years of remission. However, proving this hypothesis requires *in vitro* and *in vivo* experiments where a mixture of genotypes is exposed to the appropriate therapeutic pressure.

It is also noteworthy that patients relapsing with mutations activating signaling pathways had significantly shorter OS compared to patients relapsing with CNAs (Figure 3.18a). This suggests that: (a) signaling mutations may provide a more potent fitness advantage in this therapeutic context, warranting careful clinical monitoring of these somatic lesions; and/or (b) CNAs may hinder cellular proliferation to a certain extent. Rather surprisingly, however, not all clones bearing signaling mutations expanded at relapse (*NRAS* in pt9; *BRAF* in pt20; *CBL* in pt4; and *PTPN11* in pt10). This was unexpected, given that some of these mutations were previously associated with resistance to ivosidenib and venetoclax [305, 462, 463]. These apparent discrepancies may again be related to the clonal context: co-occurring mutations, as well as alternative competing clones, may influence the relative fitness advantage conferred by a single mutation. Future cohorts will clarify whether these preliminary observations hold true in a larger patient population.

6.2.4 Implications for clinical practice

Despite the limited cohort size, results presented here hold important implications for clinical practice and future randomized Phase III clinical trials. If my observations are reproduced in larger cohorts, this would: (a) corroborate the prognostic value of early clonal assessment specifically within LSCs, rather than bulk BM MNCs,

which may limit the ability to detect resistant clones; (b) complement current MRD methods; and (c) enable potential early clinically actionable endpoints. Detailed clonal analysis of LSC is not routinely performed in clinical practice. Yet, my results suggests that such screening may be informative of patient outcome, as clonal expansion even within a single small BM population may lead to relapse. MRD is currently assessed via immunophenotypic or genetic markers. Whilst MRD status has prognostic value [260–263], future studies will corroborate whether granular genetic screening of LSCs has superior value for prognostic purpouses. My results suggest that in several cases, despite most of the BM being devoid of resistance-associated mutations at CR, selection within the appropriate population, regardless of its small size, dictates outcome. Moreover, given that, in some patients, resistant clones became detectable only during therapy exposure, an unbiased mutational screening rapidly after initiating treatment, currently not part of standard practice, may be beneficial. This may enable to switch therapeutic strategy early, subject to ethical and practical considerations. These concepts are likely translatable to other treatments, as each therapy provides a selective pressure favoring the outgrowth of contextually fit clones.

6.2.5 Challenges ahead and future outlook

One critical question remains unresolved: why did resistant clones not emerge in the 2 patients who stayed in durable remission? It is tempting to speculate that, in all patients who relapse, resistant clones are already present at diagnosis, albeit often undetected due to their small size. In support of this, prior studies have shown that relapse-originating clones, rather than emerging *de novo*, are usually present at very low frequency already at baseline [313, 465]. Hence, in most cases, clonal selection, rather than induction, provides the basis for relapse. In this regard, the 2 patients attaining sustained remission may have never harbored clones with relapsing potential. If this hypothesis was true, it would indicate that the pre-existing clonal milieu defines the risk of relapse, and that ultra-deep sequencing at trial entry may potentially improve patient risk stratification. Ultimately, more

research into fundamental mutagenic processes, which underly the accumulation of somatic mutations, is required to devise more efficient prevention strategies. Moreover, investigation of factors driving variable clonal richness and evenness between patients is warranted to elucidate the substrate for the Darwinian selection of acquired resistance.

6.3 GTAC: integrating clonal identity with chromatin accessibility

The second part of this thesis was devoted to the development, validation, and application of GTAC, a novel multi-modal method for joint single-cell genotyping and chromatin accessibility profiling. Below, I discuss the main technical advances provided by GTAC, the key novel findings, and potential improvements and limitations of this method.

6.3.1 Technical improvement of GTAC over prior approaches

Chapter 4 demonstrated the feasibility and performance of GTAC, showing efficient simultaneous high-confidence genotyping of multiple nuclear genomic loci with high-fidelity open chromatin readout. Prior technologies have efficiently integrated single-cell genotyping with either gene expression [80, 251, 253, 324, 325, 327, 329, 331], DNA methylation [326] or surface proteome readouts [275, 328]. However, these methods were unable to link clonal identity with the landscape of accessible chromatin, precluding to study how mutations mechanistically alter gene regulation, TF activity, and epigenetic state. Whilst clonal identity was linked to epigenetic state via mitochondrial genotyping [384–386], these approaches did not allow to study how specific driver somatic mutations alter chromatin, due to the reliance on mitochondrial variants for clonal tracking. This motivated the development of GTAC.

I have shown that single-cell genotyping, which is required for capture of mutant genomic loci, enables accurate mutation detection and zygosity estimation, with low ADO rates. Multiple mutations can be detected from the same single cells,

allowing to identify clones within complex clonal hierarchies. In parallel, scATAC-seq libraries yield high complexity, signal-to-noise ratio, and minimal contamination. Due to protocol optimizations, GTAC libraries displayed improved metrics over prior published approaches, and higher cell throughput compared to the standard plate-based scATAC-seq [349], owing to an improved indexing strategy.

In parallel with GTAC, in 2024, a similar approach was developed for 10X Chromium, termed GoTChA [388]. An obvious advantage of GoTChA is the higher cell throughput. However, single-cell genotyping in GoTChA suffers from significantly higher ADO rates, leading to potential confounding effects and dilution of real differences between genotypes. Moreover, whilst genotyping in GTAC is independent of chromatin accessibility, genotyping in GoTChA partially relies on locus accessibility, causing uneven genotyping across cell populations. Finally, scATAC-seq library complexity is higher in GTAC compared to 10X Chromium, although these figures may change with time, due to optimizations to droplet-based platforms. Overall, whilst GotChA is more widely accessible and less expensive, GTAC is uniquely suitable for the study of somatically complex disease.

Linking somatic mutations and chromatin accessibility enables several advantages over prior methods coupling single-cell genotyping to RNA-seq. First, chromatin provides more granularity in distinguishing cell populations, and is overall a stronger predictor of cell identity [92]. Secondly, GTAC enables to study how mutations perturb TF networks, by providing clone-specific readouts of binding motif accessibility. Perhaps most importantly, studying how specific peaks of open chromatin vary in different clones enables the generation of novel mechanistic insights into how mutations perturb gene regulation via cis-regulatory elements.

6.3.2 Insights into *RUNX1*-associated leukaemic transformation

Applying GTAC to primary AML uncovered important insights into epigenetic evolution downstream of gradual mutation acquisition. In a first instance, this analysis confirmed that pre-leukaemic mutations in *TET2* and *ASXL1* do not

drastically alter differentiation landscapes. In contrast, the biallelic *RUNX1*-R201G LoF mutation was the key driver event associated with full-blown leukaemic transformation, suggesting strong transforming potential of *RUNX1* mutations. Despite the recurrence of *RUNX1* lesions in myeloid malignancy, how *RUNX1* mutations support leukaemogenesis is unclear. In adult haematopoiesis, RUNX1 is necessary for differentiation of immature lympho-myeloid progenitors (section 1.1.3. The current understanding is that loss of RUNX1 function perturbs this process, altering lineage decisions within immature progenitors. Consistently, *RUNX1*-mutant AMLs display a characteristic expression profile, with up-regulation of lymphoid genes [394, 408]. However, the mechanistic role of *RUNX1* mutations is understudied, and how *RUNX1* mutations perturb human haematopoiesis in a leukaemic setting is unknown. Compared to *RUNX1*-WT, *RUNX1*-mutant early progenitors had open chromatin substantially enriched for cell cycle and stress signatures, indicating a highly proliferative state. Whilst *RUNX1*-mutant progenitors showed lower activity of stemness-associated TFs, several TFs important for AML maintenance were more active in *RUNX1*-mutant cells, thus pinpointing several putative drivers of *RUNX1*-mutant leukaemia to be further investigated. Importantly, the LMPP was the earliest stage at which *RUNX1* deficiency exhibited notable consequences. Whilst exquisitely small in a healthy BM [90], the LMPP compartment was expanded in this patient, and most LMPPs were *RUNX1*-mutant. In this regard, the *RUNX1* mutation may either directly promote the transition to a highly proliferative LMPP-like state or, alternatively, may prevent LMPPs to properly execute lineage decisions, causing their abnormal expansion. Of note, most of these cells were homozygous mutant at the *RUNX1* locus, and I cannot exclude that heterozygous *RUNX1* mutations may have less pronounced consequences. These analyses also shed light on the normal role of RUNX1 in haematopoietic differentiation, suggesting the LMPP may be the earliest progenitor stage at which RUNX1 exhibits its function. GTAC analysis now needs to be performed on a higher number of *RUNX1*-mutant patients to understand whether these findings hold true in a larger cohort. This may ultimately pave the way towards new targeted therapies

for *RUNX1*-mutant AML. Of note, I cannot, at present, discern what the impact of earlier *TET2* and *ASXL1* mutations was for *RUNX1*-associated transformation; indeed, it may be that the transforming potential of *RUNX1* mutations varies according to co-occurring mutations.

6.3.3 The impact of somatic mutations on LSC gene regulation

Further epigenetic consequences of 2 subclonal *BCOR* mutations within *RUNX1*-mutant leukaemic progenitors were then studied. In AML, LoF *BCOR* mutations confer adverse prognosis [172]. Mounting evidence indicates a role of these mutations for aberrant reactivation of stemness programs in myeloid progenitors. In mice, conditional *Bcor* deletion within HSCs expands immature myeloid progenitors, associated with aberrant expression of HSC-specific genes [424]. Recent data has also suggested that loss of PRC1.1 (of which BCOR is the key component) leads to de-repression of stemness-associated genes in *NUP98*-r leukaemic cells [469]. In AML, *BCOR* and *RUNX1* frequently co-occur [296, 297, 423]; *BCOR* mutations are also frequently acquired in myeloid malignancies arising from the *RUNX1* familial platelet disorder [470]. Recent data from primary human $CD34^+$ HSPCs with co-deletion of *RUNX1* and *BCOR* showed that *BCOR* loss partially rescues stem cell dysfunction induced by *RUNX1* mutations, *in vitro* [319]. Altogether, *BCOR* loss alters normal progenitor differentiation, which seems to be particularly beneficial for *RUNX1*-mutated cells.

Whilst *BCOR*-mutant and *BCOR*-WT clones were intermingled within the leukaemic hierarchy, GTAC allowed to distinguish these clones within the same leukaemic populations. This revealed that *BCOR*-mutant clones accumulated at a primitive differentiation stage resembling LSCs and had decreased differentiation capacity. Clone-specific chromatin accessibility within molecular LSCs revealed increased openness of stemness-associated gene promoters within *BCOR*-mutant cells. This was accompanied by increased activity of several TF families involved in LSC and HSC maintenance, as the HOXA/B family, in *BCOR*-mutant LSCs. These analyses identified epigenetic de-repression of the *INPP4B* gene as a potential

therapeutic vulnerability of *BCOR*-mutant LSCs. Several lines of evidence point to a role of *INPP4B* in maintaining stemness phenotypes [231, 434, 438, 439]. The INPP4B enzyme promotes lysosome biogenesis, a process highly relevant for HSCs self-renewal, as asymmetric lysosomal inheritance by HSC daughter cells regulates their fate decisions [23] (the daughter cell inheriting a higher lysosomal content is less likely to differentiate). This may also explain the recent findings that *INPP4B* is important for LSC self-renewal [434]. Moreover, *INPP4B* is likely to be positively regulated by RUNX1 [88, 448], which may in part account for the decreased self-renewal potential of *RUNX1*-deficient cells [151, 319]. Accordingly, my results suggest a hypothesis where *BCOR* mutations increase the fitness of *RUNX1*-deficient cells, by de-repressing *INPP4B*. This, in turn, may confer increased stemness to *RUNX1*/*BCOR*-mutant LSCs. However, these preliminary findings need to be interpreted with caution. First, results derive from a single patient, and GTAC analysis of additional *RUNX1*/*BCOR*-mutant patients is required. Nevertheless, it is noteworthy that *INPP4B* was one of the few genes with altered accessibility in both *BCOR*-mutant clones. Second, GTAC does not provide definitive evidence that *INPP4B* is highly expressed in *BCOR*-mutant clones. In this regard, methods integrating scRNA-seq and genotyping may be helpful. Finally, validation experiments consisting of *INPP4B* silencing in *RUNX1*/*BCOR*-mutant cells are required to demonstrate functional relevance of *INPP4B* for these cells. Given that *INPP4B* is an enzyme, and, as such, potentially druggable, these studies may uncover therapeutic vulnerabilities for this patient subgroup.

Another interesting observation was that *BCOR*/*RUNX1*-mutant LSCs share a previously unappreciated molecular resemblance to *MLL*r or *NPM1*-mutant leukaemic cells. These leukaemias typically upregulate stemness-associated genes, like the *HOXA* cluster, *PBX3*, *MEF2C*, *JMJD1C*, *SOCS2* [201, 208, 211, 444, 446, 471]. Menin inhibitors recently showed promise in treatment of *MLL*r or *NPM1*-mutant AML patients [213–216]. The fact that *BCOR*/*RUNX1*-mutant LSCs molecularly resemble *MLL*r/*NPM1*-mutant cells suggests these cells may be sensitive to menin inhibition. However, recent findings suggest that loss of

PRC1.1 confers resistance to menin inhibitors [469]; whilst supporting the notion that PRC1.1 regulates *MLL*r-linked programs, these findings also imply that menin inhibitors alone may not be sufficient to drive differentiation of PRC1.1-deficient cells, and combinatorial strategies may be of benefit.

Interestingly, *BCOR* mutations are frequently associated with acquired resistance to mutant IDH inhibition [302] or chemoresistance [472]. Given that high stemness signatures also correlate with resistance to mutant IDH inhibition [302], this provides further evidence that *BCOR* mutations promote stemness phenotypes and are selected within a context favouring stem-like cells. My analyses uncover several putative downstream targets whose dysregulation may mediate these effects.

Finally, direct comparison of the two *BCOR*-mutant subclones revealed an unexpected epigenetic difference between clones with similar genotype. In fact, the larger clone displayed stronger epigenetic de-repression of stemness-associated genes and a more profound differentiation delay. These differences may arise due to several reasons: (a) different COOs; (b) undetected genetic differences between the two clones (non-coding mutations, for example); (c) the two *BCOR* mutations may not be functionally equivalent (the mutation in TARB2 occurs precisely in the region encoding the BCL-6 interaction domain, whilst this region is spared in TARB1); or (d) epigenetic drift or divergent epigenetic evolution. In support of the first hypothesis, prior studies have shown that the introduction of the same genetic alteration in cells at distinct differentiation stages or molecular states leads to distinct consequences [320, 473–475]. GTAC thus highlighted an additional layer of complexity in human malignancy, resulting from epigenetic variation between genetically similar clonal entities. Nevertheless, the fact that some genes, as *INPP4B*, were de-regulated in both clones, raises the hopes that common mutation-associated therapeutic vulnerabilities may exist.

6.3.4 Future applications of GTAC

Beyond AML, GTAC could be used to study any human malignancy. In fact, prior studies have demonstrated that scATAC-seq yields robust results across a wide

spectrum of human neoplastic tissues [476], suggesting GTAC is widely applicable for cancer genomics. Of note, solid tumors often harbour more somatic mutations compared to AML [173]. In this study, I rigorously showed that 7 genomic loci can be efficiently genotyped in a primary human sample. Moreover, preliminary data suggests that scATAC-seq libraries are not disrupted even when genotyping primers for 20 loci are used. However, due to costs and time required for the protocol, genotyping hundreds of loci with GTAC is unfeasible. Hence, analysis of genetically complex malignancies will be possible only if a tractable number of pre-selected mutations is genotyped (for example, the putatively most important genetic drivers).

GTAC may also be used to study how mutations alter epigenetic regulation under different external pressures. For example, future application on longitudinal patient samples will uncover how treatments induce epigenetic changes within resistant genetic clones. Importantly, results from Chapter 3 have shown that therapy resistance can arise from non-genetic clonal adaptation to treatment (Figures 3.12, 3.13, 3.14). One possible explanation for these findings is therapy-induced epigenetic evolution. In this regard, GTAC may uncover changes in the regulatory genome allowing clones to bypass treatments.

Moreover, GTAC may shed light on how genetic and epigenetic alterations cooperatively disrupt oncogenes and tumor suppressors. Whilst both categories of genes are frequently mutated in cancer cells, epigenetic silencing of tumor suppressors, or aberrant activation of oncogenes, may cooperate with the mutational landscape present within the cell. For example, *FLT3*, one of the most frequently mutated genes in AML, was not mutated in the analysed AML patient, but leukaemic cells showed increased chromatin accessibility at the *FLT3* locus. Hence, even if not mutated, *FLT3* was epigenetically primed for high expression. Analysis of a larger number of samples may thus elucidate recurrent patterns of cooperativity between genetic and epigenetic aberrations involving key driver genes.

Finally, GTAC may dissect the molecular mechanisms underlying clonal advantage in pre-malignant somatic mosaicism, such as CH. The most frequently mutated

genes in human CH are *DNMT3A*, *TET2*, and *ASXL1*, all of which epigenetic regulators. It is thus to be expected that somatic mutations in these genes may perturb cis-regulatory elements and TF activity. Studying chromatin variations downstream of these mutations in primary human samples may lead to mechanistic insights into clonal advantage induced by DTA mutations.

6.3.5 Technical considerations, limitations, and future outlook

Several limitations exist in GTAC. Firstly, due to plate-based processing, cell throughput is lower compared to droplet-based methods [388]. Hence, generating data from hundreds of thousands of cells is unfeasible. This can be an important limiting factor when studying very small clones, as GTAC may not yield enough cells for high-confidence analysis of clone-specific chromatin differences. Yet, studying small clones may be highly relevant, as their chromatin landscape may prime them for growth years before they reach a substantial size. Second, GTAC does not allow to discover somatic mutations *de novo*, as genotyping primer panels are informed by prior targeted NGS or WES. This may be an issue if there is limited sample or if GTAC needs to be performed on fresh sample. Third, although GTAC can provide information about genome regulation downstream of mutations, caution needs to be applied to avoid data over-interpretation. For example, high gene accessibility does not necessarily correspond to high gene expression, as other factors, like the RNA polymerase, histone modifications, the superelongation complex activity, and 3D genome structure, can actively influence transcription. In this regard, capturing genotype, chromatin accessibility, and RNA from the same single cells would be beneficial. Although such method has still not been developed, recent approaches for joint scATAC-seq and scRNA-seq, including the plate-based ISSAAC-seq [477], make integration of targeted genotyping theoretically feasible.

Appendices

Chapter A

Appendices

Contents

A.1 An Overview of Targeted Therapies in Acute Myeloid Leukemia 201
A.2 Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging 212
A.3 GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells 266
A.4 A protocol for simultaneous high-sensitivity genotyping and chromatin accessibility profiling in single cells 312
A.5 Mutations identified by bulk targeted panels and WES at baseline (related to Chapter 3) 366
A.6 Mutations identified by bulk targeted panels and WES at relapse (related to Chapter 3) 367
A.7 Pre-amplification primers used for TARGET-seq+ (related to Chapter 3) 368
A.8 Nested (PCR1) primers used for TARGET-seq+ (related to Chapter 3) 369
A.9 Rates of successful clonal assignment per sample (related to Chapter 3) 370
A.10 inferCNV genome-wide for pt15 (related to Chapter 3) 371
A.11 inferCNV genome-wide for pt5 (related to Chapter 3) 372
A.12 inferCNV genome-wide for pt14 (related to Chapter 3) 373
A.13 inferCNV genome-wide for pt9 (related to Chapter 3) 374
A.14 inferCNV genome-wide for pt4 (related to Chapter 3) 375
A.15 Inferred clone sizes within all live, Lin$^-$ cells for pt11 (related to Chapter 3) 376
A.16 Inferred clone sizes within all live, Lin$^-$ cells for pt5 (related to Chapter 3) 377
A.17 Inferred clone sizes within all live, Lin$^-$ cells for pt14 (related to Chapter 3) 378
A.18 Inferred clone sizes within all live, Lin$^-$ cells for pt9 (related to Chapter 3) 379
A.19 Inferred clone sizes within all live, Lin$^-$ cells for pt4 (related to Chapter 3) 380
A.20 Inferred clone sizes within all live, Lin$^-$ cells for pt18 (related to Chapter 3) 381
A.21 Inferred clone sizes within all live, Lin$^-$ cells for pt10 (related to Chapter 3) 382

A.22 Single-cell genotyping for pt11 (related to Chapter 3) 383
A.23 Single-cell genotyping for pt15 (related to Chapter 3) 384
A.24 Single-cell genotyping for pt20 (related to Chapter 3) 385
A.25 Single-cell genotyping for pt5 (related to Chapter 3) . 386
A.26 Single-cell genotyping for pt14 (related to Chapter 3) 387
A.27 Single-cell genotyping for pt9 (related to Chapter 3) . 388
A.28 Single-cell genotyping for pt4 (related to Chapter 3) . 389
A.29 Single-cell genotyping for pt18 (related to Chapter 3) 390
A.30 Single-cell genotyping for pt10 (related to Chapter 3) 391
A.31 Genotyping primers used for GTAC (related to Chapters 4 and 5) 392

A.1 An Overview of Targeted Therapies in Acute Myeloid Leukemia

pdfpages

HemaSphere

EHA EUROPEAN HEMATOLOGY ASSOCIATION

Review Article
Open Access

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

An Overview of Targeted Therapies in Acute Myeloid Leukemia

Sven Turkalj[1,2], Felix A. Radtke[1,2,3], Paresh Vyas[1,2,4]

Correspondence: Paresh Vyas (paresh.vyas@imm.ox.ac.uk).

ABSTRACT

Acute myeloid leukemia (AML) is the most aggressive adult leukemia, characterized by clonal differentiation arrest of progenitor or precursor hematopoietic cells. Intense preclinical and clinical research has led to regulatory approval of several targeted therapeutics, administered either as single agents or as combination therapies. However, the majority of patients still face a poor prognosis and disease relapse frequently occurs due to selection of therapy-resistant clones. Hence, more effective novel therapies, most likely as innovative, rational combination therapies, are urgently needed. Chromosomal aberrations, gene mutations, and epigenetic alterations drive AML pathogenesis but concurrently provide vulnerabilities to specifically target leukemic cells. Other molecules, either aberrantly active and/or overexpressed in leukemic stem cells, may also be leveraged for therapeutic benefit. This concise review of targeted therapies for AML treatment, which are either approved or are being actively investigated in clinical trials or recent preclinical studies, provides a flavor of the direction of travel, but also highlights the current challenges in AML treatment.

Acute myeloid leukemia (AML) is the most aggressive adult acute leukemia, which still has a high unmet clinical need. Although treatment outcomes have improved, and multiple new drugs have been licensed over the last few decades, most patients still face a dismal prognosis. Approximately 70% of AML patients are above the age of 65 years and are mostly treated with nonintensive chemotherapy. The current standard of care for this patient group is a combination of the BCL2 inhibitor venetoclax and the hypomethylating agent (HMA) azacitidine. The median survival is 14.5 months, and patients are usually not cured. The remaining 30% of patients, usually aged <60 years, are treated with >1 course of intensive combination chemotherapy, potentially combined with targeted therapies. Patients at high risk of treatment failure receive allogeneic stem cell transplantation (allo-SCT). Although the cure rates for this group are ~30%–70%, the majority of intensively treated patients still die of their disease.

A significant advance in AML biology and treatment has been the definition of the genetic and cytogenetic landscape of the disease.[1,2] This has provided genetic markers for diagnostic and prognostic purposes, as well as markers used to follow the depth of therapy response (also known as measurement of residual disease). In some cases, the discovery of genetic markers has led to the development of specific therapies targeting mutant proteins essential for AML propagation.

Therapeutic targets in AML can be broadly classified into 6 categories (Table 1; Figure 1).

FUSION PROTEINS: THE PARADIGM FOR TARGETED THERAPY IN MYELOID MALIGNANCIES

Recurrent chromosomal aberrations can generate fusion proteins, which cause differentiation block and sustain malignant transformation, but may also create a critical therapeutic vulnerability of leukemic cells.[3] The fusion protein stemming from the Philadelphia chromosome in chronic myelogenous leukemia (CML),[4] which generates the BCR-ABL1 oncogene with constitutive tyrosine kinase activity, is the exemplar of a targetable fusion protein in myeloid malignancy. Imatinib,[5] a specific inhibitor of the BCR-ABL1 kinase domain,[6] transformed the care of CML patients and introduced the concept of targeting products of somatic alteration in cancer. Targeted therapies have also revolutionized the treatment of acute promyelocytic leukemia (APL), caused by the promyelocytic leukemia-retinoic acid receptor alpha (PML-RARA) fusion oncoprotein,[7,8] which blocks granulocytic differentiation.[9] Patients now routinely receive all-trans retinoic acid (ATRA), which reverses transcriptional repression imposed by PML-RARA, thus promoting granulocyte differentiation. ATRA is combined with arsenic trioxide (ATO), which degrades PML-RARA, removing the oncogenic protein that sustains APL. A chemotherapy-free regimen combining ATRA and ATO has led to >90% overall survival (OS).[10,11] This approach highlighted the potential for differentiation therapy in leukemia[12] and demonstrated the importance of combining targeted therapies with synergistic mechanisms of action.[13]

[1]MRC Molecular Hematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, United Kingdom
[2]Oxford Centre for Hematology, NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom
[3]Department of Medicine V, Hematology, Oncology and Rheumatology, Heidelberg University Hospital, Heidelberg, Germany
[4]Department of Hematology, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
ST and FAR have contributed equally to this work.
Copyright © 2023 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.
HemaSphere (2023) 7:6(e914).
http://dx.doi.org/10.1097/HS9.0000000000000914.
Received: April 18, 2023 / Accepted: May 1, 2023

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

The therapeutic success in CML and APL has been harder to replicate for other oncogenic fusion proteins, especially those involving transcription factors. One approach to developing targeted therapy for AML driven by transcription factor fusion proteins is to target pathways downstream of these transcription factors. An example is the inhibition of Menin,[14] a required coactivator for MLL (KMT2A) fusion proteins. Menin and oncogenic MLL fusion proteins coordinately activate the transcription of *HOX* family members and *MEIS1*.[15,16] These observations, coupled with preclinical data showing promising activity of Menin inhibition,[17–19] prompted first-in-human clinical trials of multiple Menin inhibitors, namely revumenib,[20,21] ziftomenib,[22] and JNJ-6617 (NCT04811560), in *MLL*-rearranged and *NPM1*-mutated leukemia. As one may expect, second-site mutations in the *MEN1* gene, encoding Menin, confer resistance to revumenib.[21] MLL fusions also interact with the histone methyltransferase DOT1L, required for aberrant activation of target genes.[23] In a phase I study, the DOT1L inhibitor pinometostat showed limited single-agent activity in *MLL*-rearranged adult AML.[24] Pinometostat is now being evaluated in combination with azacitidine.[25]

Table 1

Therapeutic Targets in AML

Type of Target	Targets
Genetic Targets	Fusion Oncoproteins, FLT3, IDH1/2, p53, KIT
Targets involved in apoptosis	BCL2, MCL-1, BCL-XL, MDM2
Signaling molecules or nuclear receptors	RAS pathway, SYK, RARα
Epigenetic regulators	Menin, DOT1L
Surface proteins and/or molecules involved in immune signaling	Smoothened, CD33, CD123, CD47, SIRPα, AXL, TIM-3, CLL-1, PD-1/PD-L1, and CTLA-4 axes
Transcription factors and other molecules subject to targeted protein degradation	EZH2, GSPT1

GENETIC TARGETS RESULTING FROM SOMATIC MUTATION AND OTHER ABERRANTLY ACTIVE TARGETS

Driver somatic mutations, which produce mutated proteins imparting a fitness advantage to AML cells, also provide an opportunity to develop targeted therapies (Table 2). For example, gain-of-function mutations in the *FLT3* gene, encoding a transmembrane receptor tyrosine kinase (RTK), are found in ~30% of younger AML patients and at lower frequency in older patients. Most *FLT3* mutations occur as internal tandem duplications (*FLT3*-ITD) within the gene locus encoding the juxtamembrane domain or as point mutations within the locus encoding the tyrosine kinase domain activation loop (*FLT3*-TK). Mutant FLT3 proteins transmit a constitutively active ligand-independent signal to aberrantly activate prosurvival and proproliferative RAS/MAPK, PI3K/AKT, and JAK/STAT pathways.[26] Type I FLT3 inhibitors, such as midostaurin, gilteritinib, and crenolanib, bind the kinase active conformation, inhibiting both FLT3-ITD and FLT3-TK proteins. Type II FLT3 inhibitors, including

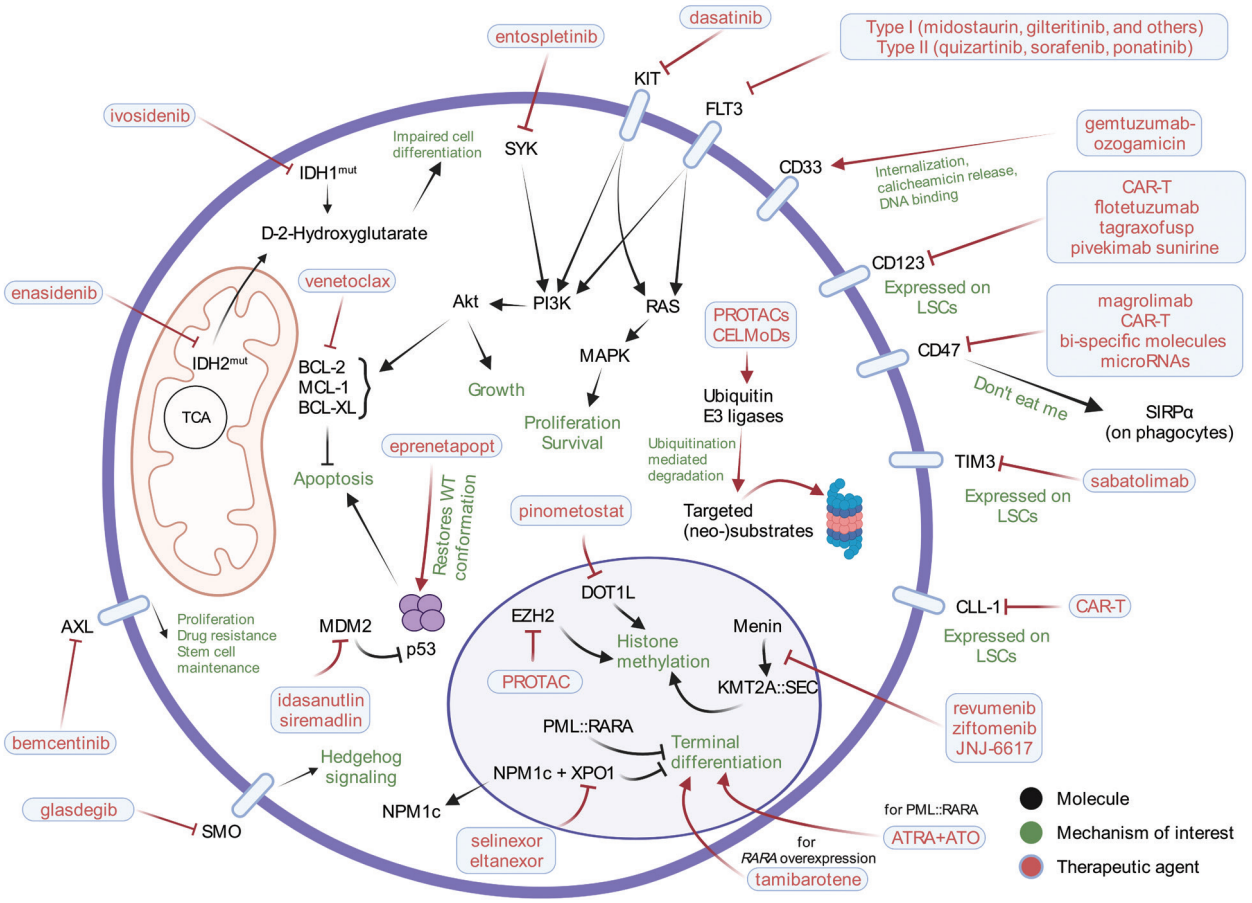


Figure 1. Overview of therapeutic agents and targeted pathways in AML. Schematic overview of targeted molecules found either on the AML cell surface or within the cytoplasm, nucleus, or mitochondria of AML cells. The molecules used as therapeutic agents are labeled in red. Targeted molecules are labeled in black. Cellular mechanisms involved or targeted molecule expression patterns are labeled in green. AML = acute myeloid leukemia.

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

Table 2

Overview of Targeted Therapies in Acute Myeloid Leukemia, Which are Either Approved or Are Being Investigated in Clinical Trials

Target	Therapeutic(s)	Mechanism of Action
PML-RARA	All-trans retinoic acid + arsenic trioxide	PML-RARA configuration change, which results in coregulator exchange; sumoylation and degradation of PML-RARA
Menin	Revumenib	Menin-MLL interaction inhibition
	Ziftomenib	Menin-MLL interaction inhibition
	JNJ-6617	Menin-MLL interaction inhibition
DOT1L	Pinometostat	Specific DOT1L inhibition via conformational change
FLT3	Type I (targeting FLT3-ITD and FLT3-TK) Midostaurin Crenolanib Gilteritinib	FLT3 inhibition, binds FLT3 receptor in active conformation
	Type II (targeting FLT3-ITD) Quizartinib Sorafenib	FLT3 inhibition, binds FLT3 receptor in inactive conformation
Several kinases (KIT, SRC family, BCR-ABL, others)	Dasatinib	ATP-competitive protein tyrosine kinase inhibition
Mutated IDH1	Ivosidenib	Selective allosteric inhibition of mutant IDH1
Mutated IDH2	Enasidenib	Selective allosteric inhibition of mutant IDH2
Mutant p53	Eprenetapopt	Covalent p53 core domain binding, partial restoration of WT function
XPO1	Selinexor	Selective XPO1 inhibition
	Eltanexor	Selective XPO1 inhibition
SYK	Entospletinib	Selective SYK inhibition
RARα	Tamibarotene	RARα agonism
BCL2	Venetoclax	BCL2 sequestration (BH3-mimetic)
MCL-1	Murizatoclax	Selective MCL-1 inhibition
MEK	Cobimetinib	Selective MEK1 inhibition
MDM2	Idasanutlin	p53-MDM2 interaction inhibition
	Siremadlin	p53-MDM2 interaction inhibition
Smoothened	Glasdegib	Inhibition of the translocation of Smoothened to the primary cilium
CD33	Gemtuzumab ozogamicin	Gets internalized, subsequent calicheamicin causes DNA double-strand breaks
CD123	Tagraxofusp	Anti-CD123 antibody fused to diphtheria toxin, gets internalized and blocks protein synthesis via eEF2 inhibition
	Pivekimab sunirine	Anti-CD123 antibody fused to indolinobenzodiazepine pseudodimer payload, leading to DNA alkylation and single-strand breaks
CD3/CD123	Flotetuzumab, UCART123v1.2	Anti-CD3/CD123 bispecific antibody; recruitment of T cells to CD123-expressing cells
CLL-1	CAR-T cells	Recruitment of T cells to CLL-1-expressing cells
TIM-3	Sabatolimab	Anti-TIM-3 monoclonal antibody—blockage of ligand binding
CD47/SIRPα	Magrolimab, TTI-621	Inhibition of the CD47-SIRPα axis
AXL	Bemcentinib	Selective AXL inhibition
CTLA-4	Ipilimumab	Anti-CTLA-4 monoclonal antibody—immune checkpoint inhibition
PD-1	Nivolumab, pembrolizumab	Anti-PD-1 monoclonal antibody—immune checkpoint inhibition
GSPT1	CC-90009	Linkage of GSPT1 to an E3 ubiquitin ligase—promotes GSPT1 degradation

quizartinib and sorafenib, bind the kinase-inactive conformation and are only active against FLT3-ITD. These inhibitors decrease FLT3 autophosphorylation, reducing FLT3 signaling.

Following successful phase III clinical trials, midostaurin[27] was approved in combination with chemotherapy in frontline *FLT3*-mutated AML, while gilteritinib was approved as single-agent in *FLT3*-mutated relapsed or refractory AML (R/R AML).[28] In addition, results of a phase III trial of quizartinib combined with intensive chemotherapy showed superior OS compared with intensive chemotherapy alone,[29] and a decision on regulatory approval is awaited. A common mechanism of resistance to FLT3 inhibitors arises from secondary mutations in the ATP-binding region within the FLT3 kinase domain,[30,31] which seem to be inhibitor-specific.[31] Alternatively, activating mutations in genes encoding members of downstream RAS/MAPK or JAK/STAT pathways, or persistent activation of pathways downstream of FLT3 due to stimulation from microenvironmental factors, can be selected by treatment and mediate resistance.[32–36] Accordingly, preclinical data indicate that combining FLT3 and MEK inhibition could be more efficient in causing leukemic cell death.[34,37] FLT3 inhibitors are actively being evaluated in combination with venetoclax,[38] venetoclax/azacitidine,[39] venetoclax/ASTX727,[40] and as single agents post allo-SCT[41,42] (NCT02997202) or postchemotherapy[43] (NCT02668653 for quizartinib with chemotherapy or as continuation therapy).

Activating mutations in *KIT*, encoding another RTK upstream of PI3K/AKT, RAS/MAPK, and JAK/STAT pathways, occur in a subset of core-binding factor (CBF) AML patients. Notably, KIT D816 mutations have been associated with poorer outcomes in chemotherapy-treated patients.[44] Dasatinib is a multikinase inhibitor active against KIT, SRC family, BCR-ABL, and other kinases. Phase II clinical trials of dasatinib with intensive chemotherapy reported favorable event-free survival and cumulative incidence of relapse in newly diagnosed CBF AML patients.[45,46] Notably, survival outcomes were compared favorably with a historical control group treated with chemotherapy alone.[45] Ongoing adequately powered phase III trials will better elucidate dasatinib's role in this setting (NCT02013648).

Approximately 20% of AMLs harbor hotspot gain-of-function mutations in *IDH1* and *IDH2*, encoding isocitrate dehydrogenase (IDH) enzymes.[47] Wild-type IDH catalyzes the conversion of isocitrate to α-ketoglutarate (α-KG) in the tricarboxylic acid (TCA) cycle. Neomorphic mutant IDH catalyzes the production of the oncometabolite 2-hydroxyglutarate (2-HG)[48] from α-KG, resulting in competitive inhibition of α-KG-dependent enzymes, including ten-eleven translocation (TET) family of 5-methylcytosine hydroxylases[49] and Jumonji-C domain histone demethylases.[50,51] This leads to DNA hypermethylation and impaired hematopoietic differentiation.[52] The small molecules ivosidenib and enasidenib, allosteric inhibitors of mutant IDH1 and IDH2, respectively, reverse DNA hypermethylation and promote differentiation.[47] Both agents were approved by the FDA for the treatment of *IDH*-mutated AML.[53–55] Notably, cooccurring *RUNX1* and *NRAS* mutations, as well as a more stem-like transcriptional state of leukemic cells, correlated with poorer responses to mutant IDH inhibition and may mediate primary resistance.[56,57] Unsurprisingly, when mutant IDH inhibitors are used as monotherapy, disease relapse occurs by selecting therapy-resistant clones that have acquired multiple genetic resistance mechanisms.[56,58] These include *TET2*, *BCOR*, *FLT3*, and RAS pathway mutations, as well as mutations in transcription factors involved in hematopoietic differentiation, which may re-establish differentiation block.[56,58] Restoration of 2-HG through either secondary *IDH* mutations,[59,60] which compromise inhibitor binding, or via selection of mutations on the gene encoding the second IDH isoform (isoform switching),[58,59,61] have also been reported. Consequently, both ivosidenib and

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

enasidenib are being tested in combination with chemotherapy, HMAs, and venetoclax.[53,62–64]

TP53 mutations occur in 8%–12% of AML patients, a significant proportion of which also have complex or monosomal karyotypes.[65] These genetic features correlate with universally poor prognosis.[2,66] The small molecule eprenetapopt covalently binds to the core domain of p53,[67] which mediates the cooperative DNA binding of p53. In some cases, this leads to appropriate refolding and restoration of p53 function. Eprenetapopt demonstrated encouraging activity in early-phase trials when combined with azacitidine[68,69] and when administered post allo-SCT,[70] but this activity was not validated in a phase III trial in *TP53*-mutated higher-risk MDS, which failed to meet its primary end point of a higher complete remission (CR) rate (NCT03745716).

Approximately, 30% of AML patients harbor hotspot mutations in exon 12 of the *NPM1* gene,[2,71] causing a frameshift that forms a novel C-terminal nuclear export signal (NES) and causes loss of a nucleolar localization signal. This results in the NPM1c neomorphic oncoprotein with predominantly cytoplasmic localization.[72,73] Due to the novel C-terminal NES, NPM1c is more efficiently exported from the nucleus and binds with higher affinity to the nuclear exporter XPO1 (CRM1).[74] It has been proposed that increased nuclear export of NPM1c hijacks several proteins from the nucleus, including PU.1[75] and CTCF,[76] thus disrupting their normal function.[74] Furthermore, NPM1c interferes with the p53 pathway by destabilizing the Arf tumor suppressor.[77] Notably, *NPM1*-mutated AML typically displays a high expression of *HOXA*/*HOXB* cluster and *MEIS1* genes,[78] which depends on NPM1c.[79] In this regard, 2 recent studies have demonstrated that the nuclear portion of NPM1c directly regulates the transcription of several target genes, including *HOX* family and *MEIS1*.[80,81] In fact, NPM1c binds chromatin through the interaction with XPO1 and cooperates with MLL/Menin activity, likely underscoring favorable activity of Menin and DOT1L inhibitors in *NPM1*-mutated AML models.[82,83] Notably, combined inhibition of XPO1 and Menin showed a synergistic decrease in *HOX* and *MEIS1* expression, decreased colony-forming capacity of AML blasts, and induction of differentiation in model systems.[80,81] The XPO1 inhibitor selinexor has shown promise in early trials on patients with hematologic malignancies, including AML,[84] while the next-generation XPO1 inhibitor eltanexor demonstrated reduced toxicity compared with selinexor in preclinical models.[85] Hence, there is a sound rationale for testing the combination of XPO1 and Menin inhibitors in *NPM1*-mutated AML patients.

Other molecules, which are aberrantly expressed and/or active, rather than mutant, have also been targeted therapeutically. One exemplar is the non-RTK SYK that has been shown to cooperate both with FLT3-ITD[86] and HOXA9/MEIS1.[87] In a phase Ib/II study in previously untreated AML patients, the SYK inhibitor entospletinib combined with induction chemotherapy led to composite CR of 70%, with potentially improved OS in patients with high basal *HOXA9* and *MEIS1* expression.[88] A phase II trial of entospletinib combined with induction and consolidation chemotherapy in newly diagnosed *NPM1*-mutated/*FLT3*-ITD-wild-type AML patients reported a composite CR rate of 78% and a median duration of response of 15.7 months.[89] A phase III trial of entospletinib with chemotherapy in previously untreated *NPM1*-mutated AML patients is ongoing (NCT05020665).

Interestingly, the *RARA* gene is overexpressed in a significant proportion of non-APL AML and MDS patients. This has provided the impetus for the development of the RARα agonist tamibarotene. In preclinical models of RARα-high AML, tamibarotene induced AML differentiation, reminiscent of ATRA activity in APL.[90] A phase II clinical trial combining tamibarotene and azacitidine showed that 61% of evaluable RARα-high patients achieved CR or CR with incomplete hematological recovery (CRi).[91] Ongoing trials are testing tamibarotene with venetoclax/azacitidine in newly diagnosed AML or with azacitidine in higher-risk myelodysplastic syndrome (MDS) (NCT04905407 and NCT04797780).

TARGETS INVOLVED IN APOPTOSIS

Regulation of apoptosis is often aberrant in cancer.[92] A significant proportion of AML cells are highly dependent on the antiapoptotic BCL2 protein for their survival.[93,94] Furthermore, BCL2 is overexpressed in functionally defined leukemia stem cells (LSCs).[95] Concordantly, high levels of BCL2 correlate with chemoresistance in AML.[96] The BH3-mimetic venetoclax is a selective BCL2 inhibitor that sequesters BCL2, thus precluding its binding to proapoptotic proteins. Venetoclax demonstrated efficacy when combined with azacitidine or low-dose cytarabine (LDAC) in treatment-naïve AML patients.[97,98] These combinations are now the standard of care for AML patients unfit for intensive chemotherapy.

Importantly, response to venetoclax, or venetoclax combinations, varies with the genetic subtype of AML. *IDH1*/2-mutated AML has high response rates and more durable remissions to venetoclax.[99] These clinical observations are concordant with preclinical data showing that *IDH*-mutated AML cells are highly dependent on BCL2 for survival due to the inhibition of cytochrome c oxidase by 2-HG, which decreases the threshold for apoptosis.[100] Of note, remission rates for *IDH2*-mutated patients treated with venetoclax/azacitidine were particularly high, even when compared with *IDH1*-mutated patients.[101] For *IDH1*-mutated AML, venetoclax has also recently been combined with ivosidenib and ivosidenib/azacitidine, resulting in encouraging preliminary response rates.[63] Poorer responses to venetoclax/azacitidine have been noted with *FLT3*– (in the absence of *NPM1* mutation), *TP53*–, and *RAS*-mutated AML.[102,103] Interestingly, venetoclax combined with the protein translation inhibitor omacetaxine showed encouraging preclinical activity in *RUNX1*-mutated AML models.[104]

Despite these promising results, multiple mechanisms of resistance to venetoclax have been described in AML.[105] For example, the selection of the monocytic differentiation program[106] is associated with increased expression and function of the antiapoptotic protein MCL-1. Consistent with MCL-1 mediating venetoclax resistance, several preclinical studies have shown that inhibition of BCL2 and MCL-1 synergize in AML models.[107–109] Clinical trials with MCL-1 inhibitors have been ongoing for some time but have been complicated by on-target cardiac toxicity (NCT03465540). An additional mechanism of venetoclax resistance includes the selection of clones with mutations in *FLT3*, *TP53*, and genes mediating RAS/MAPK signalling.[105] Hence, to potentially mitigate resistance, rational combinations of venetoclax or venetoclax/azacitidine with inhibitors of FLT3,[38,39,105,110] MEK,[111,112] or MDM2[113,114] are being actively studied in preclinical and clinical settings. By combining clonal tracking and RNA-Seq at single-cell resolution, we are currently investigating the mechanisms of therapy response and resistance to venetoclax and ivosidenib +/– azacitidine.[63]

Besides induction of apoptosis, BCL2 inhibition also decreases levels of oxidative phosphorylation (OxPhos) in functionally defined LSCs.[95] Moreover, the venetoclax/azacitidine treatment disrupts amino acid uptake[115] and the TCA cycle,[116] with consequent decrease in OxPhos and energy metabolism in these cells. Consistently, resistance to venetoclax/azacitidine arises from upregulation of fatty acid oxidation, which replenishes energy production,[117] and from increased nicotinamide levels,[118] which promote both fatty acid oxidation and amino acid metabolism, thus driving OxPhos. Hence, variations in central energy metabolism represent an important resistance mechanism to venetoclax/azacitidine and correlate with mutations in *TP53*[119] and genes mediating RAS/MAPK signaling.[117] Importantly, MCL-1

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

inhibition decreased fatty acid oxidation and engraftment of venetoclax/azacitidine-resistant LSCs.[117] These studies not only further highlight the potential benefits of MCL-1 inhibition in venetoclax/azacitidine-resistant AML but also encourage clinical investigation of agents targeting intermediate metabolism, with key considerations regarding treatment specificity.

SURFACE PROTEIN TARGETING

Several therapeutic strategies target proteins expressed on the AML cell surface (Table 2). One target is the smoothened transmembrane protein, which mediates Hedgehog signaling by activating GLI transcription factors in proximity to the cellular structure known as primary cilium. Glasdegib inhibits the translocation of smoothened to the primary cilium, thus blunting GLI activation.[120] Although there is limited knowledge about the exact role of this pathway in AML, encouraging data from a phase II clinical trial showed that the addition of glasdegib to LDAC improved median OS compared with LDAC alone (8.8 months versus 4.9 months), in older or unfit AML and high-risk MDS patients.[121] This led to the licensing of a smoothened inhibitor glasdegib combined with LDAC in the USA (2018) and in the EU (2020). A current phase III trial is evaluating glasdegib in combination with intensive chemotherapy or azacitidine in previously untreated AML patients (NCT03416179).

CD33 and CD123 molecules are generally found on subsets of hematopoietic cells but are highly expressed on the AML cell surface in most patients.[122] Gemtuzumab ozogamicin is a humanized anti-CD33 monoclonal antibody conjugated to the cytotoxic drug calicheamicin. Upon binding to CD33, gemtuzumab ozogamicin is internalized, and calicheamicin is released, causing DNA double-strand breaks, which trigger cell death. Gemtuzumab ozogamicin has been approved in combination with intensive chemotherapy for newly diagnosed AML.[123,124] Interestingly, gemtuzumab ozogamicin, in combination with intensive chemotherapy, has maximal benefit in good prognosis CBF AML, but the reasons for this are still not entirely clear.

Tagraxofusp is a recombinant anti-CD123 (IL-3 receptor alpha chain) antibody fused to diphtheria toxin. When bound to CD123, tagraxofusp is internalized and blocks protein synthesis via eEF2 inhibition. Tagraxofusp has shown clinical benefit in blastic plasmacytoid dendritic cell neoplasm, a specific form of myeloid malignancy with high CD123 expression.[125] Tagraxofusp has also been combined with azacitidine or venetoclax/azacitidine in CD123-positive AML and MDS patients. Eight of the 9 previously untreated AML patients and 3 of 4 of previously untreated MDS patients achieved CR/CRi.[126] Of note, 2 responding AML patients and 3 responding MDS patients were *TP53*-mutated. Another anti-CD123 antibody-drug conjugate, pivekimab sunirine, which causes DNA alkylation,[127] led to objective response rates (ORR) of 51% when combined with venetoclax/azacitidine in CD123-positive R/R AML, with auspicious activity in the *FLT3*-ITD subgroup, where ORR was 82%.[128]

Bispecific antibodies, typically bispecific T-cell engagers, recruit T cells to tumors via an anti-CD3 moiety by binding to a tumor-specific antigen. An exemplar is blinatumomab, used to treat B-cell acute lymphoblastic leukemia. The success of blinatumomab has triggered enormous activity in this therapeutic area with variations in the mechanisms of action. One example is the concept of dual affinity retargeting antibodies such as flotetuzumab, an anti-CD3/CD123 bispecific antibody.[129] In phase 1 of 2 studies, 26.6% R/R AML patients who did not respond to induction therapy or who experienced early relapse postinduction therapy achieved CR/CRh with a median OS of 10.2 months. Interestingly, a 10-gene signature derived from diagnostic bone marrow was predictive of CR.

Chimeric antigen receptor (CAR) T-cell therapy has been transformative in lymphoid malignancy. CD19-directed CAR-T led the initial wave of therapies,[130–132] which was recognized by the American Society of Clinical Oncology[133] as the advance of the year in 2018. Between 2018 and 2023, 5 CAR-T cell therapies targeting CD19 or B-cell maturation antigen were authorized for sale. However, finding a suitable antigen target for AML that is not present on normal myeloid cells has been a challenge. CD123-directed CAR-T cells showed promising antileukemic activity in vitro and in vivo, with limited toxicity toward healthy hematopoietic cells.[134] A phase I trial evaluated UCART123v1.2, an allogeneic anti-CD123 T-cell product, in 16 CD123-positive R/R AML patients.[135] Four patients showed evidence of activity, including 1 long-term CR. Another approach has been universal CAR-T therapy with inducible activity of the CD123-targeting module. In a phase I trial of 14 CD123-positive R/R AML, 3 patients achieved CR/CRi, with 4 partial responses.[136] Anti-CD123 CAR-T cells have also been tested in pediatric R/R AML patients[137] in a phase I study. CAR-T cells against the surface molecule CLL-1 have shown some promise, with 7 of 10 R/R AML patients achieving CR/CRi in a phase I trial.[138]

TIM-3 is a surface receptor preferentially expressed on AML cells with LSC function compared with normal HSCs.[139] Moreover, TIM-3 marks dysfunctional and/or exhausted subsets of T cells in viral infections and cancer.[140,141] Concordantly, xenograft data in nonobese diabetic-severe combined immunodeficiency (NOD-SCID) mice showed a decrease in AML reconstitution and LSC function following antibody targeting of TIM-3.[142] Sabatolimab, an anti-TIM-3 monoclonal antibody, combined with HMAs, achieved a CR/CRi rate of 30% in the Revised International Prognostic Scoring System for high-/very high-risk MDS and newly diagnosed AML patients unfit for intensive chemotherapy.[143] More extensive phase II/III clinical trials are currently evaluating the efficacy of sabatolimab in multiple regimens.[144]

An alternative targeted strategy harnesses autologous innate immunity to specifically kill leukemic cells. CD47 (integrin-associated protein) is a surface molecule mediating *don't eat me* signals. This occurs via CD47-SIRPα interactions, which inhibit macrophage-mediated phagocytosis.[145] Notably, CD47 is upregulated both on LSC-enriched and nonenriched AML cells, compared with normal hematopoietic cells, predicting poor outcome.[146,147] Targeting the CD47-SIRPα axis has thus led to multiple drugs in preclinical and clinical development, including monoclonal antibodies, fusion proteins, bispecific antibodies, small molecules, RNA interference, and CD47-directed CAR-T cells. Importantly, anti-CD47 monoclonal antibodies stimulated AML cell phagocytosis in vivo while sparing normal HSCs.[146] This encouraged the development of magrolimab, a humanized anti-CD47 antibody, which was well-tolerated in R/R AML patients in the CAMELLIA phase 1 trial.[148] Results from a subsequent phase IB trial combining magrolimab with azacitidine in 95 untreated intermediate/high- and very high-risk MDS patients have just been reported. Sixty-two percent of patients had poor-risk cytogenetics, and 26% had a *TP53* mutation.[149] The combination was well-tolerated, the CR rate was 33%, and the duration of the overall response was 9.8 months. Median OS was not reached with a median 17.1-month follow-up. About 40% of *TP53*-mutated patients achieved CR and had a median OS of 16.1 months. Results from a phase Ib trial of magrolimab+azacitidine in 87 previously untreated AML patients, 72 of which *TP53*-mutated (median variant allele frequency of 61%), have also just been reported (Daver et al, invited revision resubmitted at *Journal of Clinical Oncology*). About 79.2% of *TP53*-mutated patients had adverse-risk cytogenetics according to the European LeukemiaNet (ELN) 2017 classification. Totally 28 (32.2%) patients achieved CR, including 23 (31.9%) patients with *TP53* mutations. Median OS in *TP53*-mutated patients was 9.8 months and 18.9 months in *TP53*-wild-type patients. Magrolimab has also been combined with venetoclax/azacitidine in a phase I/II study, displaying an ORR of 80% in newly diagnosed AML and 75% in venetoclax-naïve R/R AML.[150] Of note, 63% of newly diagnosed *TP53*-mutant patients achieved

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

CR/CRi. Three phase III trials of magrolimab combinations in previously untreated higher-risk MDS and AML patients are in progress and will determine its clinical role (NCT04313881, NCT05079230, and NCT04778397). Other approaches targeting the CD47-SIRPα axis include the fusion protein TTI-621 (SIRPα-IgG1 Fc), which inhibited the growth of AML xenografts,[151] leading to its evaluation in early-phase clinical trials for hematological malignancies.[152] In addition, the bispecific molecule PD1xCD47 BsAb HX009 has very recently shown antileukemic activity in a patient-derived xenograft model of AML.[153]

MicroRNAs repress gene expression by selective binding to the 3′ untranslated regions of target genes, which induces mRNA degradation and inhibition of translation.[154] In a preclinical study, miR-708 was shown to negatively regulate *CD47* expression and to inhibit T-ALL engraftment.[155] These alternative approaches might thus hold promise in reactivating antileukemic autologous innate immunity.

AXL is a member of the TAM family of RTK exerting pleiotropic effects in cancer, including the potentiation of migratory capabilities.[156] *Axl* overexpression correlates with poor prognosis in AML,[157] and *Axl* expression is an independent prognostic factor in de novo AML.[158] Bemcentinib is a selective small molecule AXL inhibitor, which is well-tolerated as single agent by AML patients who relapsed after treatment with HMAs.[159] In an ongoing phase II trial, bemcentinib is being investigated in AML patients either as single agent or in combination with LDAC or decitabine.[160]

While immune checkpoint inhibition has shown great promise in several solid tumors, results in AML have been disappointing after some initial encouraging data with ipilimumab (CTLA-4 inhibition) in relapsed patients post allo-SCT.[161] Since then, CTLA-4 (ipilimumab) and PD-1 (nivolumab and pembrolizumab) inhibitors have shown modest efficacy in phase I and II trials[161–169] and when combined with azacitidine in R/R AML.[170] Furthermore, PD-1 inhibition combined with chemotherapy showed some efficacy in newly diagnosed AML,[167] and PD-1 inhibition combined with venetoclax-azacitidine is being evaluated (NCT04284787).

TARGETED PROTEIN DEGRADATION

The cellular protein degradation machinery can be therapeutically leveraged to degrade proteins that cancer cells depend on (Table 2). The ubiquitin/proteasome system, composed of E3 ubiquitin ligase complexes, ubiquitinates damaged and unwanted proteins, directing them toward proteasomal degradation. In this context, the therapeutic agent acts as molecular glue, linking E3 ubiquitin ligases to target proteins (termed neosubstrates). For example, the small molecules thalidomide, lenalidomide, and pomalidomide link neosubstrates to cereblon, a member of the CRL4 E3 ubiquitin ligase complex. This leads to the degradation of proteins essential for multiple myeloma or MDS cell survival.[171] Subsequent development of Cereblon E3 ligase modulating drugs (CELMoDs) allowed the targeting of new neosubstrates by linking them to cereblon-containing E3 ligases. One CELMoD being investigated in AML is CC-90009[172]: This molecule promotes the degradation of GSPT1, which coregulates the termination of protein translation. In preclinical studies, GSPT1 degradation inhibited AML cell survival and proliferation.[173] Of note, CC-90009 efficacy may depend on AML differentiation stage.[174] CC-90009, as a single agent and in combination with other targeted approaches, is being evaluated in phase I trials in AML patients (NCT02848001 and NCT04336982). We are currently investigating clonal and molecular determinants of response and resistance to this treatment.

One limitation of the above approaches is the requirement for neosubstrates to contain a common *degron* peptide recognized by cereblon. Proteolysis-targeting chimeras (PROTACs) bind both to the target protein (via a PROTAC warhead) and to an E3 ubiquitin ligase enzyme via a linker domain connected to the E3 ligand.[3] This allows degrading proteins without the requirement for a common *degron* peptide. PROTACs are being investigated in hematological malignancies, for example, to target BCL2 in chronic lymphocytic leukemia.[175] The PROTAC technology has recently been designed to specifically degrade EZH2, the core member of the PRC2 complex with histone methyltransferase activity.[176] Mounting evidence suggests that EZH2 inhibits differentiation and promotes AML maintenance.[177–179] A recent study further identified a noncatalytic role of EZH2 in MLL-rearranged AML caused by the direct binding of EZH2 to cMYC and p300, which promotes malignant cell growth.[176] The novel PROTAC MS177, which promotes EZH2 degradation via cereblon, disrupted both the PRC2-dependent and cMYC-dependent EZH2 activity, inhibiting leukemic cell growth in vitro and in vivo.[176] Hence, this study exemplifies how PROTACs may be particularly effective in abolishing additional noncatalytic functions of targeted proteins. PROTACs might prove effective against transcription factors with oncogenic function in AML, such as WT1, MEIS1, or HOX family members.

CHALLENGES AHEAD: HOW TO OVERCOME THEM?

Novel targeted therapies are revolutionizing AML therapy, but fundamental challenges persist. The selection of resistant clones under therapy pressure causes treatment failure and relapse. An integrated characterization of genetic and molecular landscapes in longitudinal patient samples is needed to better understand putative resistance mechanisms. Emerging single-cell approaches coupling somatic genotypes with cellular readouts are potent tools to describe clonal complexity.[180] Complementing these technologies, our group recently integrated clonal readout with chromatin accessibility profiling at single-cell resolution, linking genetic and epigenetic disease evolution.[181] Going forward, this approach will elucidate clone-specific epigenetic mechanisms underlying disease relapse.

An obstacle to cure is the failure to eradicate quiescent LSCs, the seed of malignant growth most frequently residing in leukemic progenitor and/or precursor populations.[182,183] Hence, a better understanding of mechanisms driving aberrant self-renewal of LSCs is necessary and might lead to novel therapeutic targets.

In the context of disease prevention, the study of preleukemic clonal outgrowths, occurring years before leukemia onset, is critical. We and others are actively exploring molecular landscapes underlying human clonal hematopoiesis,[184–186] to elucidate pathways leading to early clonal dominance predating myeloid malignancy.

Going forward, better coordination between basic, translational, and clinical research is critical to maximize the value of technological advances. This will not only boost research through improved sample availability but also will lead to the rapid implementation of new technologies in clinical settings. Larger multicenter trials are warranted for the efficient delivery of promising therapeutic agents. Moreover, the process of drug discovery and clinical evaluation will benefit from increased collaboration between governments, industries, trial centers, and academic research. Finally, huge steps are needed to improve science communication between patients, their families, researchers, clinicians, and the public, to keep the primary focus on improving patients' lives.

ACKNOWLEDGMENTS

Figure 1 was created using BioRender. The authors thank Dr. Maria Chiara Barbanti and Dr. Bing Tseu for their insightful comments and discussions.

AUTHOR CONTRIBUTIONS

ST and FAR did writing and original draft preparation. PV did writing, review, and editing. All authors have read and agreed to the final version of the article.

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

DISCLOSURES

The authors have no conflicts of interest to disclose.

SOURCES OF FUNDING

PV acknowledges funding from the Medical Research Council Molecular Hematology Unit Programme Grant (MC_UU_00029/8), Blood Cancer UK Programme Continuity Grant 13008, NIHR Senior Fellowship, and the Oxford BRC Hematology Theme. ST is supported by a Scatcherd European Scholarship in partnership with The Medical Research Council/Radcliffe Department of Medicine and The Clarendon Fund. FAR is supported by the Oxford – Sir David Weatherall Scholarship from Green Templeton College and the WIMM Prize Studentship from the MRC Weatherall Institute of Molecular Medicine.

REFERENCES

1. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult De Novo acute myeloid leukemia. *N Engl J Med*. 2013;368:2059–2074.
2. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. 2016;374:2209–2221.
3. Pikman Y, Stegmaier K. Targeted therapy for fusion-driven high-risk acute leukemia. *Blood*. 2018;132:1241–1247.
4. Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and giemsa staining. *Nature*. 1973;243:290–293.
5. Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med*. 2001;344:1031–1037.
6. Druker BJ, Tamura S, Buchdunger E, et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr–Abl positive cells. *Nat Med*. 1996;2:561–566.
7. de Thé H, Chomienne C, Lanotte M, et al. The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor α gene to a novel transcribed locus. *Nature*. 1990;347:558–561.
8. Kakizuka A, Miller WH Jr, Umesono K, et al. Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML. *Cell*. 1991;66:663–674.
9. Wang Z-Y, Chen Z. Acute promyelocytic leukemia: from highly fatal to highly curable. *Blood*. 2008;111:2505–2515.
10. Lo-Coco F, Avvisati G, Vignetti M, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med*. 2013;369:111–121.
11. Hu J, Liu Y-F, Wu C-F, et al. Long-term efficacy and safety of all-trans retinoic acid/arsenic trioxide-based therapy in newly diagnosed acute promyelocytic leukemia. *Proc Natl Acad Sci USA*. 2009;106:3342–3347.
12. Meng-er H, Yu-chen Y, Shu-rong C, et al. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood*. 1988;72:567–572.
13. Lallemand-Breitenbach V, Guillemin M-C, Janin A, et al. Retinoic acid and arsenic synergize to eradicate leukemic cells in a mouse model of acute promyelocytic leukemia. *J Exp Med*. 1999;189:1043–1052.
14. Yokoyama A, Somervaille TCP, Smith KS, et al. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*. 2005;123:207–218.
15. Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. *Genes Dev*. 2011;25:661–672.
16. Soto-Feliciano YM, Sanchez-Rivera FJ, Perner F, et al. A molecular switch between mammalian MLL complexes dictates response to menin-MLL inhibition. *Cancer Discov*. 2023;13:146–169.
17. Borkin D, He S, Miao H, et al. Pharmacologic inhibition of the menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell*. 2015;27:589–602.
18. Fiskus W, Daver N, Boettcher S, et al. Activity of menin inhibitor ziftomenib (KO-539) as monotherapy or in combinations against AML cells with MLL1 rearrangement or mutant NPM1. *Leukemia*. 2022;36:2729–2733.
19. Fiskus W, Boettcher S, Daver N, et al. Effective Menin inhibitor-based combinations against AML with MLL rearrangement or NPM1 mutation (NPM1c). *Blood Cancer J*. 2022;12:5.
20. Issa GC, Aldoss I, DiPersio J, et al. The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia. *Nature*. 2023;615:920–924.
21. Perner F, Stein EM, Wenge DV, et al. MEN1 mutations mediate clinical resistance to menin inhibition. *Nature*. 2023;615:913–919.
22. Erba HP, Fathi AT, Issa GC, et al. Update on a phase 1/2 first-in-human study of the menin-KMT2A (MLL) inhibitor ziftomenib (KO-539) in patients with relapsed or refractory acute myeloid leukemia. *Blood*. 2022;140(Supplement 1):153–156.
23. Bernt Kathrin M, Zhu N, Sinha Amit U, et al. MLL-Rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell*. 2011;20:66–78.
24. Stein EM, Garcia-Manero G, Rizzieri DA, et al. The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood*. 2018;131:2661–2669.
25. Menghrajani K, Cai SF, Devlin SM, et al. A phase Ib/II study of the histone methyltransferase inhibitor pinometostat in combination with azacitidine in patients with 11q23-rearranged acute myeloid leukemia. *Blood*. 2019;134(Supplement_1):2655–2655.
26. Daver N, Schlenk RF, Russell NH, et al. Targeting FLT3 mutations in AML: review of current knowledge and evidence. *Leukemia*. 2019;33:299–312.
27. Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med*. 2017;377:454–464.
28. Perl AE, Martinelli G, Cortes JE, et al. Gilteritinib or chemotherapy for relapsed or refractory FLT3-mutated AML. *N Engl J Med*. 2019;381:1728–1740.
29. Erba H, Montesinos P, Vrhovac R, et al. S100: quizartinib prolonged survival vs placebo plus intensive induction and consolidation therapy followed by single-agent continuation in patients aged 18-75 years with newly diagnosed FLT3-ITD+ AML. *HemaSphere*. 2022;6:1–2.
30. Heidel F, Solem FK, Breitenbuecher F, et al. Clinical resistance to the kinase inhibitor PKC412 in acute myeloid leukemia by mutation of Asn-676 in the FLT3 tyrosine kinase domain. *Blood*. 2006;107:293–300.
31. von Bubnoff N, Engh RA, Åberg E, et al. FMS-like tyrosine kinase 3–internal tandem duplication tyrosine kinase inhibitors display a nonoverlapping profile of resistance mutations in vitro. *Cancer Res*. 2009;69:3032–3041.
32. Piloto O, Wright M, Brown P, et al. Prolonged exposure to FLT3 inhibitors leads to resistance via activation of parallel signaling pathways. *Blood*. 2006;109:1643–1652.
33. Alotaibi AS, Yilmaz M, Kanagal-Shamanna R, et al. Patterns of resistance differ in patients with acute myeloid leukemia treated with type I versus type II FLT3 inhibitors. *Blood Cancer Discov*. 2021;2:125–134.
34. McMahon CM, Ferng T, Canaani J, et al. Clonal selection with RAS pathway activation mediates secondary clinical resistance to selective FLT3 inhibition in acute myeloid leukemia. *Cancer Discov*. 2019;9:1050–1063.
35. Traer E, Martinez J, Javidi-Sharifi N, et al. FGF2 from marrow microenvironment promotes resistance to FLT3 inhibitors in acute myeloid leukemia. *Cancer Res*. 2016;76:6471–6482.
36. Rummelt C, Gorantla SP, Meggendorfer M, et al. Activating JAK-mutations confer resistance to FLT3 kinase inhibitors in FLT3-ITD positive AML in vitro and in vivo. *Leukemia*. 2021;35:2017–2029.
37. Joshi SK, Nechiporuk T, Bottomly D, et al. The AML microenvironment catalyzes a stepwise evolution to gilteritinib resistance. *Cancer Cell*. 2021;39:999–1014.e8.
38. Daver N, Perl AE, Maly J, et al. Venetoclax plus gilteritinib for FLT3-mutated relapsed/refractory acute myeloid leukemia. *J Clin Oncol*. 2022;40:4048–4059.
39. Short N, DiNardo CD, Daver N, et al. Updated results from a phase I/II study of the triplet combination of azacitidine, venetoclax and gilteritinib for patients with FLT3-mutated acute myeloid leukemia. *Blood*. 2022;140(Supplement 1):2007–2009.
40. Ong F, Short N, Daver N, et al. A phase I/II study of combination of ASTX727, gilteritinib and venetoclax in patients with relapsed/refractory FLT3 mutated acute myeloid leukemia (AML) and frontline FLT3 mutated AML patients unfit for chemotherapy. *Blood*. 2022;140(Supplement 1):9029–9031.
41. Xuan L, Wang Y, Huang F, et al. Sorafenib maintenance in patients with FLT3-ITD acute myeloid leukaemia undergoing allogeneic haematopoietic stem-cell transplantation: an open-label, multicentre, randomised phase 3 trial. *Lancet Oncol*. 2020;21:1201–1212.
42. Perl AE, Larson RA, Podoltsev NA, et al. Follow-up of patients with R/R FLT3-mutation–positive AML treated with gilteritinib in the phase 3 ADMIRAL trial. *Blood*. 2022;139:3366–3375.
43. Schlenk RF, Weber D, Fiedler W, et al. Midostaurin added to chemotherapy and continued single-agent maintenance therapy in acute myeloid leukemia with FLT3-ITD. *Blood*. 2019;133:840–851.

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

44. Cairoli R, Beghini A, Grillo G, et al. Prognostic impact of c-KIT mutations in core binding factor leukemias: an Italian retrospective study. *Blood*. 2006;107:3463–3468.
45. Paschka P, Schlenk RF, Weber D, et al. Adding dasatinib to intensive treatment in core-binding factor acute myeloid leukemia—results of the AMLSG 11-08 trial. *Leukemia*. 2018;32:1621–1630.
46. Marcucci G, Geyer S, Laumann K, et al. Combination of dasatinib with chemotherapy in previously untreated core binding factor acute myeloid leukemia: CALGB 10801. *Blood Adv*. 2020;4:696–705.
47. Issa GC, DiNardo CD. Acute myeloid leukemia with IDH1 and IDH2 mutations: 2021 treatment algorithm. *Blood Cancer J*. 2021;11:107.
48. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 2009;462:739–744.
49. Xu W, Yang H, Liu Y, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell*. 2011;19:17–30.
50. Chowdhury R, Yeoh KK, Tian Y-M, et al. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep*. 2011;12:463–469.
51. Lu C, Ward PS, Kapoor GS, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*. 2012;483:474–478.
52. Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010;18:553–567.
53. Kantarjian H, Short NJ, DiNardo C, et al. Harnessing the benefits of available targeted therapies in acute myeloid leukaemia. *Lancet Haematol*. 2021;8:e922–e933.
54. DiNardo CD, Stein EM, de Botton S, et al. Durable remissions with ivosidenib in IDH1-mutated relapsed or refractory AML. *N Engl J Med*. 2018;378:2386–2398.
55. Stein EM, DiNardo C, Altman JK, et al. Safety and efficacy of AG-221, a Potent inhibitor of mutant IDH2 that promotes differentiation of myeloid cells in patients with advanced hematologic malignancies: results of a phase 1/2 trial. *Blood*. 2015;126:323–323.
56. Wang F, Morita K, DiNardo CD, et al. Leukemia stemness and co-occurring mutations drive resistance to IDH inhibitors in acute myeloid leukemia. *Nat Commun*. 2021;12:2607.
57. Amatangelo MD, Quek L, Shih A, et al. Enasidenib induces acute myeloid leukemia cell differentiation to promote clinical response. *Blood*. 2017;130:732–741.
58. Quek L, David MD, Kennedy A, et al. Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib. *Nat Med*. 2018;24:1167–1177.
59. Choe S, Wang H, DiNardo CD, et al. Molecular mechanisms mediating relapse following ivosidenib monotherapy in IDH1-mutant relapsed or refractory AML. *Blood Adv*. 2020;4:1894–1905.
60. Intlekofer AM, Shih AH, Wang B, et al. Acquired resistance to IDH inhibition through trans or cis dimer-interface mutations. *Nature*. 2018;559:125–129.
61. Harding JJ, Lowery MA, Shih AH, et al. Isoform switching as a mechanism of acquired resistance to mutant isocitrate dehydrogenase inhibition. *Cancer Discov*. 2018;8:1540–1547.
62. DiNardo CD, Stein AS, Stein EM, et al. Mutant isocitrate dehydrogenase 1 inhibitor ivosidenib in combination with azacitidine for newly diagnosed acute myeloid leukemia. *J Clin Oncol*. 2021;39:57–65.
63. Lachowiez CA, Loghavi S, Zeng Z, et al. A phase Ib/II study of ivosidenib with venetoclax +/- azacitidine in IDH1-mutated myeloid malignancies. *Blood Cancer Discov*. 2023 April 27. [Epub ahead of print].
64. Stein EM, DiNardo CD, Fathi AT, et al. Ivosidenib or enasidenib combined with intensive chemotherapy in patients with newly diagnosed AML: a phase 1 study. *Blood*. 2021;137:1792–1803.
65. Bullinger L, Döhner K, Döhner H. Genomics of acute myeloid leukemia diagnosis and pathways. *J Clin Oncol*. 2017;35:934–946.
66. Rücker FG, Schlenk RF, Bullinger L, et al. TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. *Blood*. 2012;119:2114–2121.
67. Lambert JMR, Gorzov P, Veprintsev DB, et al. PRIMA-1 reactivates mutant p53 by covalent binding to the core domain. *Cancer Cell*. 2009;15:376–388.
68. Cluzeau T, Sebert M, Rahmé R, et al. Eprenetapopt plus azacitidine in TP53-mutated myelodysplastic syndromes and acute myeloid leukemia: a phase II Study by the Groupe Francophone des Myélodysplasies (GFM). *J Clin Oncol*. 2021;39:1575–1583.
69. Sallman DA, DeZern AE, Garcia-Manero G, et al. Eprenetapopt (APR-246) and azacitidine in TP53-mutant myelodysplastic syndromes. *J Clin Oncol*. 2021;39:1584–1594.
70. Mishra A, Tamari R, DeZern AE, et al. Eprenetapopt plus azacitidine after allogeneic hematopoietic stem-cell transplantation for TP53-mutant acute myeloid leukemia and myelodysplastic syndromes. *J Clin Oncol*. 2022;40:3985–3993.
71. Falini B, Nicoletti I, Martelli MF, et al. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood*. 2006;109:874–885.
72. Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med*. 2005;352:254–266.
73. Falini B, Bolli N, Shan J, et al. Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. *Blood*. 2006;107:4514–4523.
74. Falini B, Brunetti L, Sportoletti P, et al. NPM1-mutated acute myeloid leukemia: from bench to bedside. *Blood*. 2020;136:1707–1721.
75. Gu X, Ebrahem Q, Mahfouz RZ, et al. Leukemogenic nucleophosmin mutation disrupts the transcription factor hub that regulates granulomonocytic fates. *J Clin Invest*. 2018;128:4260–4279.
76. Wang AJ, Han Y, Jia N, et al. NPM1c impedes CTCF functions through cytoplasmic mislocalization in acute myeloid leukemia. *Leukemia*. 2020;34:1278–1290.
77. Colombo E, Martinelli P, Zamponi R, et al. Delocalization and destabilization of the arf tumor suppressor by the leukemia-associated NPM mutant. *Cancer Res*. 2006;66:3044–3050.
78. Alcalay M, Tiacci E, Bergomas R, et al. Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood*. 2005;106:899–902.
79. Brunetti L, Gundry MC, Sorcini D, et al. Mutant NPM1 maintains the leukemic state through HOX expression. *Cancer Cell*. 2018;34:499–512.e9.
80. Uckelmann HJ, Haarer EL, Takeda R, et al. Mutant NPM1 directly regulates oncogenic transcription in acute myeloid leukemia. *Cancer Discov*. 2023;13:746–765.
81. Wang XQD, Fan D, Han Q, et al. Mutant NPM1 hijacks transcriptional hubs to maintain pathogenic gene programs in acute myeloid leukemia. *Cancer Discov*. 2023;13:724–745.
82. Kühn MWM, Song E, Feng Z, et al. Targeting chromatin regulators inhibits leukemogenic gene expression in npm1 mutant leukemia. *Cancer Discov*. 2016;6:1166–1181.
83. Uckelmann HJ, Kim SM, Wong EM, et al. Therapeutic targeting of preleukemia cells in a mouse model of NPM1 mutant acute myeloid leukemia. *Science*. 2020;367:586–590.
84. Azmi AS, Uddin MH, Mohammad RM. The nuclear export protein XPO1 — from biology to targeted therapy. *Nat Rev Clin Oncol*. 2021;18:152–169.
85. Etchin J, Berezovskaya A, Conway AS, et al. KPT-8602, a second-generation inhibitor of XPO1-mediated nuclear export, is well tolerated and highly active against AML blasts and leukemia-initiating cells. *Leukemia*. 2017;31:143–150.
86. Puissant A, Fenouille N, Alexe G, et al. SYK is a critical regulator of FLT3 in acute myeloid leukemia. *Cancer Cell*. 2014;25:226–242.
87. Mohr S, Doebele C, Comoglio F, et al. Hoxa9 and meis1 cooperatively induce addiction to syk signaling by suppressing mir-146a in acute myeloid leukemia. *Cancer Cell*. 2017;31:549–562.e11.
88. Walker AR, Byrd JC, Blachly JS, et al. Entospletinib in combination with induction chemotherapy in previously untreated acute myeloid leukemia: response and predictive significance of HOXA9 and MEIS1 expression. *Clin Cancer Res*. 2020;26:5852–5859.
89. Borate U, Li R, Huang Y, et al. Entospletinib (ENTO) in combination with cytarabine (Ara-C) and daunorubicin (DNR) in newly diagnosed (ND) adult patients with NPM1-mutated and FLT3-ITD wild-type acute myeloid leukemia (AML) is associated with good response and survival: a phase 2 sub-study of the beat AML master trial. *Blood*. 2022;140(Supplement 1):6200–6203.
90. McKeown MR, Corces MR, Eaton ML, et al. Superenhancer analysis defines novel epigenomic subtypes of Non-APL AML, including an RARα dependency targetable by SY-1425, a potent and selective RARα agonist. *Cancer Discov*. 2017;7:1136–1153.
91. de Botton S, Cluzeau T, Vigil CE, et al. Targeting RARA overexpression with tamibarotene, a potent and selective RARα agonist, is a novel approach in AML. *Blood Adv*. 2023;7:1858–1870.
92. Singh R, Letai A, Sarosiek K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. *Nat Rev Mol Cell Biol*. 2019;20:175–193.
93. Konopleva M, Contractor R, Tsao T, et al. Mechanisms of apoptosis sensitivity and resistance to the BH3 mimetic ABT-737 in acute myeloid leukemia. *Cancer Cell*. 2006;10:375–388.
94. Vo T-T, Ryan J, Carrasco R, et al. Relative mitochondrial priming of myeloblasts and normal HSCs determines chemotherapeutic success in AML. *Cell*. 2012;151:344–355.

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

95. Lagadinou Eleni D, Sach A, Callahan K, et al. BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. *Cell Stem Cell*. 2013;12:329–341.
96. Campos L, Rouault J-P, Sabido O, et al. High expression of bcl-2 protein in acute myeloid leukemia cells is associated with poor response to chemotherapy. *Blood*. 1993;81:3091–3096.
97. DiNardo CD, Jonas BA, Pullarkat V, et al. Azacitidine and venetoclax in previously untreated acute myeloid leukemia. *N Engl J Med*. 2020;383:617–629.
98. Wei AH, Strickland SA Jr, Hou J-Z, et al. Venetoclax combined with low-dose cytarabine for previously untreated patients with acute myeloid leukemia: results from a phase Ib/II study. *J Clin Oncol*. 2019;37:1277–1284.
99. Konopleva M, Pollyea DA, Potluri J, et al. Efficacy and biological correlates of response in a phase II study of venetoclax monotherapy in patients with acute myelogenous leukemia. *Cancer Discov*. 2016;6:1106–1117.
100. Chan SM, Thomas D, Corces-Zimmerman MR, et al. Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia. *Nat Med*. 2015;21:178–184.
101. Pollyea DA, DiNardo CD, Arellano ML, et al. Impact of venetoclax and azacitidine in treatment-naïve patients with acute myeloid leukemia and IDH1/2 mutations. *Clin Cancer Res*. 2022;28:2753–2761.
102. DiNardo CD, Tiong IS, Quaglieri A, et al. Molecular patterns of response and treatment failure after frontline venetoclax combinations in older patients with AML. *Blood*. 2020;135:791–803.
103. Konopleva M, Thirman MJ, Pratz KW, et al. Impact of FLT3 mutation on outcomes after venetoclax and azacitidine for patients with treatment-naive acute myeloid leukemia. *Clin Cancer Res*. 2022;28:2744–2752.
104. Mill CP, Fiskus W, DiNardo CD, et al. Effective therapy for AML with RUNX1 mutation by cotreatment with inhibitors of protein translation and BCL2. *Blood*. 2022;139:907–921.
105. Ong F, Kim K, Konopleva MY. Venetoclax resistance: mechanistic insights and future strategies. *Cancer Drug Resist*. 2022;5:380–400.
106. Pei S, Pollyea DA, Gustafson A, et al. Monocytic subclones confer resistance to venetoclax-based therapy in patients with acute myeloid leukemia. *Cancer Discovery*. 2020;10:536–551.
107. Ramsey HE, Fischer MA, Lee T, et al. A novel MCL1 inhibitor combined with venetoclax rescues venetoclax-resistant acute myelogenous leukemia. *Cancer Discovery*. 2018;8:1566–1581.
108. Moujalled DM, Pomilio G, Ghiurau C, et al. Combining BH3-mimetics to target both BCL-2 and MCL1 has potent activity in pre-clinical models of acute myeloid leukemia. *Leukemia*. 2019;33:905–917.
109. Tron AE, Belmonte MA, Adam A, et al. Discovery of Mcl-1-specific inhibitor AZD5991 and preclinical activity in multiple myeloma and acute myeloid leukemia. *Nat Commun*. 2018;9:5341.
110. Yilmaz M, Kantarjian H, Short NJ, et al. Hypomethylating agent and venetoclax with FLT3 inhibitor "triplet" therapy in older/unfit patients with FLT3 mutated AML. *Blood Cancer J*. 2022;12:77.
111. Desikan SP, Ravandi F, Pemmaraju N, et al. A Phase II study of azacitidine, venetoclax and trametinib in relapsed/refractory AML harboring a ras pathway-activating mutation. *Blood*. 2021;138:4436–4436.
112. Padua RA, Sarda-Mantel L, Chiquet M, et al. BCL-2 inhibitor venetoclax (ABT-199) and MEK inhibitor GDC-0973 synergise to target AML progenitors and overcome drug resistance with the use of PET scanning in a mouse model of HR-MDS to monitor response to treatment. *Blood*. 2018;132(Supplement 1):5497–5497.
113. Daver NG, Dail M, Garcia JS, et al. Venetoclax and idasanutlin in relapsed/refractory AML: a nonrandomized, open-label phase 1b trial. *Blood*. 2023;141:1265–1276.
114. Wei AH, Breccia M, Ooi M, et al. Preliminary results from a phase Ib study exploring MDM2 inhibitor siremadlin (HDM201) in combination with B-Cell Lymphoma-2 (BCL-2) inhibitor venetoclax in patients with acute myeloid leukemia (AML) or high-risk myelodysplastic syndrome (HR-MDS). *Blood*. 2021;138(Supplement 1):1283–1283.
115. Jones CL, Stevens BM, D'Alessandro A, et al. Inhibition of amino acid metabolism selectively targets human leukemia stem cells. *Cancer Cell*. 2018;34:724–740.e4.
116. Pollyea DA, Stevens BM, Jones CL, et al. Venetoclax with azacitidine disrupts energy metabolism and targets leukemia stem cells in patients with acute myeloid leukemia. *Nat Med*. 2018;24:1859–1866.
117. Stevens BM, Jones CL, Pollyea DA, et al. Fatty acid metabolism underlies venetoclax resistance in acute myeloid leukemia stem cells. *Nature Cancer*. 2020;1:1176–1187.
118. Jones CL, Stevens BM, Pollyea DA, et al. Nicotinamide metabolism mediates resistance to venetoclax in relapsed acute myeloid leukemia stem cells. *Cell Stem Cell*. 2020;27:748–764.e4.
119. Nechiporuk T, Kurtz SE, Nikolova O, et al. The TP53 apoptotic network is a primary mediator of resistance to BCL2 inhibition in AML cells. *Cancer Discovery*. 2019;9:910–925.
120. Chaudhry P, Singh M, Triche TJ, et al. GLI3 repressor determines Hedgehog pathway activation and is required for response to SMO antagonist glasdegib in AML. *Blood*. 2017;129:3465–3475.
121. Cortes JE, Heidel FH, Hellmann A, et al. Randomized comparison of low dose cytarabine with or without glasdegib in patients with newly diagnosed acute myeloid leukemia or high-risk myelodysplastic syndrome. *Leukemia*. 2019;33:379–389.
122. Ehninger A, Kramer M, Röllig C, et al. Distribution and levels of cell surface expression of CD33 and CD123 in acute myeloid leukemia. *Blood Cancer J*. 2014;4:e218–e218.
123. Hills RK, Castaigne S, Appelbaum FR, et al. Addition of gemtuzumab ozogamicin to induction chemotherapy in adult patients with acute myeloid leukaemia: a meta-analysis of individual patient data from randomised controlled trials. *Lancet Oncol*. 2014;15:986–996.
124. Castaigne S, Pautas C, Terre C, et al. Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet*. 2012;379:1508–1516.
125. Pemmaraju N, Lane AA, Sweet KL, et al. Tagraxofusp in blastic plasmacytoid dendritic-cell neoplasm. *N Engl J Med*. 2019;380:1628–1637.
126. Lane AA, Stein AS, Garcia JS, et al. Safety and efficacy of combining tagraxofusp (SL-401) with azacitidine or azacitidine and venetoclax in a phase 1b study for CD123 positive AML, MDS, or BPDCN. *Blood*. 2021;138(Supplement 1):2346–2346.
127. Kovtun Y, Jones GE, Adams S, et al. A CD123-targeting antibody-drug conjugate, IMGN632, designed to eradicate AML while sparing normal bone marrow cells. *Blood Adv*. 2018;2:848–858.
128. Daver N, Montesinos P, Aribi A, et al. Broad activity for the pivekimab sunirine (PVEK, IMGN632), azacitidine, and venetoclax triplet in high-risk patients with relapsed/refractory acute myeloid leukemia (AML). *Blood*. 2022;140(Supplement 1):145–149.
129. Uy GL, Aldoss I, Foster MC, et al. Flotetuzumab as salvage immunotherapy for refractory acute myeloid leukemia. *Blood*. 2021;137:751–762.
130. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. *N Engl J Med*. 2017;377:2531–2544.
131. Maude SL, Laetsch TW, Buechner J, et al. Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. *N Engl J Med*. 2018;378:439–448.
132. Schuster SJ, Svoboda J, Chong EA, et al. Chimeric antigen receptor T cells in refractory B-Cell lymphomas. *N Engl J Med*. 2017;377:2545–2554.
133. Heymach J, Krilov L, Alberg A, et al. Clinical cancer advances 2018: annual report on progress against cancer from the American Society of Clinical Oncology. *J Clin Oncol*. 2018;36:1020–1044.
134. Sugita M, Galetto R, Zong H, et al. Allogeneic TCRαβ deficient CAR T-cells targeting CD123 in acute myeloid leukemia. *Nat Commun*. 2022;13:2227.
135. Sallman DA, DeAngelo DJ, Pemmaraju N, et al. Ameli-01: a phase I trial of UCART123v1.2, an Anti-CD123 allogeneic CAR-T cell product, in adult patients with relapsed or refractory (R/R) CD123+ acute myeloid leukemia (AML). *Blood*. 2022;140(Supplement 1):2371–2373.
136. Ehninger G, Kraus S, Sala E, et al. Phase 1 dose escalation study of the rapidly switchable universal CAR-T therapy unicar-T-CD123 in relapsed/refractory AML. *Blood*. 2022;140(Supplement 1):2367–2368.
137. Naik S, Madden RM, Lipsitt A, et al. Safety and anti-leukemic activity of CD123-CAR T cells in pediatric patients with AML: preliminary results from a phase 1 trial. *Blood*. 2022;140(Supplement 1):4584–4585.
138. Jin X, Zhang M, Sun R, et al. First-in-human phase I study of CLL-1 CAR-T cells in adults with relapsed/refractory acute myeloid leukemia. *J Hematol Oncol*. 2022;15:88.
139. Jan M, Chao MP, Cha AC, et al. Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. *Proc Natl Acad Sci USA*. 2011;108:5009–5014.
140. Jin H-T, Anderson AC, Tan WG, et al. Cooperation of Tim-3 and PD-1 in CD8 T-cell exhaustion during chronic viral infection. *Proc Natl Acad Sci USA*. 2010;107:14733–14738.
141. Sakuishi K, Apetoh L, Sullivan JM, et al. Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J Exp Med*. 2010;207:2187–2194.
142. Kikushige Y, Shima T, Takayanagi S-i, et al. TIM-3 Is a promising target to selectively kill acute myeloid leukemia stem cells. *Cell Stem Cell*. 2010;7:708–717.

Downloaded from http://journals.lww.com/hemasphere by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdgGj2MwlZLeI= on 07/31/2024

143. Brunner AM, Esteve J, Porkka K, et al. Efficacy and Safety of Sabatolimab (MBG453) in Combination with Hypomethylating Agents (HMAs) in Patients (Pts) with Very High/High-Risk Myelodysplastic Syndrome (vHR/HR-MDS) and Acute Myeloid Leukemia (AML): Final Analysis from a Phase Ib Study. *Blood*. 2021;138(Supplement 1):244–244.
144. Zeidan AM, Westermann J, Kovacsovics T, et al. P582: first results of a phase II study (STIMULUS-AML1) Investigating Sabatolimab + Azacitidine + Venetoclax In Patients With Newly Diagnosed Acute Myeloid Leukemia. *HemaSphere*. 2022;6:481–482.
145. Tsai RK, Discher DE. Inhibition of "self" engulfment through deactivation of myosin-II at the phagocytic synapse between human cells. *J Cell Biol*. 2008;180:989–1003.
146. Majeti R, Chao MP, Alizadeh AA, et al. CD47 is an adverse prognostic factor and therapeutic antibody target on human acute myeloid leukemia stem cells. *Cell*. 2009;138:286–299.
147. Jaiswal S, Jamieson CHM, Pang WW, et al. CD47 Is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell*. 2009;138:271–285.
148. Brierley CK, Staves J, Roberts C, et al. The effects of monoclonal anti-CD47 on RBCs, compatibility testing, and transfusion requirements in refractory acute myeloid leukemia. *Transfusion*. 2019;59:2248–2254.
149. Sallman DA, Al Malki MM, Asch AS, et al. Magrolimab in combination with azacitidine in patients with higher-risk myelodysplastic syndromes: final results of a phase Ib study. *J Clin Oncol*. 2023;41:2815–2826.
150. Daver N, Senapati J, Maiti A, et al. Phase I/II Study of Azacitidine (AZA) with Venetoclax (VEN) and Magrolimab (Magro) in Patients (pts) with Newly Diagnosed (ND) Older/Unfit or High-Risk Acute Myeloid Leukemia (AML) and Relapsed/Refractory (R/R) AML. *Blood*. 2022;140(Supplement 1):141–144.
151. Petrova PS, Viller NN, Wong M, et al. TTI-621 (SIRPalphaFc): a CD47-blocking innate immune checkpoint inhibitor with broad antitumor activity and minimal erythrocyte binding. *Clin Cancer Res*. 2017;23:1068–1079.
152. Ansell SM, Maris MB, Lesokhin AM, et al. Phase I study of the CD47 blocker TTI-621 in patients with relapsed or refractory hematologic malignancies. *Clin Cancer Res*. 2021;27:2190–2199.
153. An X, Liu J, Ke H, et al. Abstract 2980: HX009, a first-in-class PD1xCD47 BsAb, demonstrated anti-AML activity in PDX models. *Cancer Res*. 2023;83(7_Supplement):2980–2980.
154. Lau NC, Lim LP, Weinstein EG, et al. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*. 2001;294:858–862.
155. Huang W, Wang WT, Fang K, et al. MIR-708 promotes phagocytosis to eradicate T-ALL cells by targeting CD47. *Mol Cancer*. 2018;17:12.
156. Abu-Thuraia A, Goyette MA, Boulais J, et al. AXL confers cell migration and invasion by hijacking a PEAK1-regulated focal adhesion protein network. *Nat Commun*. 2020;11:3586.
157. Rochlitz C, Lohri A, Bacchi M, et al. Axl expression is associated with adverse prognosis and with expression of Bcl-2 and CD34 in de novo acute myeloid leukemia (AML): results from a multicenter trial of the Swiss Group for Clinical Cancer Research (SAKK). *Leukemia*. 1999;13:1352–1358.
158. Ben-Batalla I, Schultze A, Wroblewski M, et al. Axl, a prognostic and therapeutic target in acute myeloid leukemia mediates paracrine crosstalk of leukemia cells with bone marrow stroma. *Blood*. 2013;122:2443–2452.
159. Kubasch AS, Peterlin P, Cluzeau T, et al. Efficacy and Safety of Bemcentinib in Patients with Myelodysplastic Syndromes or Acute Myeloid Leukemia Failing Hypomethylating Agents. *Blood*. 2020;136(Supplement 1):37–38.
160. Loges S, Sutamtewagul G, Heuser M, et al. Bemcentinib (Oral AXL Inhibitor) in Combination with Low-Dose Cytarabine Is Well Tolerated and Efficacious in Older Relapsed AML Patients.Updates from the Ongoing Phase II Trial (NCT02488408) and Preliminary Translational Results Indicating Bemcentinib Elicits Anti-AML Immune Responses. *Blood*. 2021;138(Supplement 1):3410–3410.
161. Davids MS, Kim HT, Bachireddy P, et al. Ipilimumab for patients with relapse after allogeneic transplantation. *N Engl J Med*. 2016;375:143–153.
162. Zeidan AM, Knaus HA, Robinson TM, et al. A Multi-center phase I trial of ipilimumab in patients with myelodysplastic syndromes following hypomethylating agent failure. *Clin Cancer Res*. 2018;24:3519–3527.
163. Goswami M, Gui G, Dillon LW, et al. Pembrolizumab and decitabine for refractory or relapsed acute myeloid leukemia. *J ImmunoTher Cancer*. 2022;10:e003392.
164. Zeidner JF, Vincent BG, Ivanova A, et al. Phase II trial of pembrolizumab after high-dose cytarabine in relapsed/refractory acute myeloid leukemia. *Blood Cancer Discov*. 2021;2:616–629.
165. Tschernia NP, Kumar V, Moore DT, et al. Safety and efficacy of pembrolizumab prior to allogeneic stem cell transplantation for acute myelogenous leukemia. *Transplant Cell Ther*. 2021;27:1021 e1–1021 e5.
166. Daver N, Garcia-Manero G, Basu S, et al. Efficacy, safety, and biomarkers of response to azacitidine and nivolumab in relapsed/refractory acute myeloid leukemia: a nonrandomized, open-label, phase II study. *Cancer Discov*. 2019;9:370–383.
167. Ravandi F, Assi R, Daver N, et al. Idarubicin, cytarabine, and nivolumab in patients with newly diagnosed acute myeloid leukaemia or high-risk myelodysplastic syndrome: a single-arm, phase 2 study. *Lancet Haematol*. 2019;6:e480–e488.
168. Reville PK, Kantarjian HM, Ravandi F, et al. Nivolumab maintenance in high-risk acute myeloid leukemia patients: a single-arm, open-label, phase II study. *Blood Cancer J*. 2021;11:60.
169. Daver NG, Garcia-Manero G, Konopleva MY, et al. Azacitidine (AZA) with Nivolumab (Nivo), and AZA with Nivo + Ipilimumab (Ipi) in Relapsed/Refractory Acute Myeloid Leukemia: A Non-Randomized, Prospective, Phase 2 Study. *Blood*. 2019;134(Supplement_1):830–830.
170. Daver N, Basu S, Garcia-Manero G, et al. Azacitidine (AZA) with Nivolumab (Nivo), and AZA with Nivo + Ipilimumab (Ipi) in Relapsed/Refractory (R/R) Acute Myeloid Leukemia: Clinical and Immune Biomarkers of Response. *Blood*. 2020;136(Supplement 1):43–45.
171. Fink EC, Ebert BL. The novel mechanism of lenalidomide activity. *Blood*. 2015;126:2366–2369.
172. Surka C, Jin L, Mbong N, et al. CC-90009, a novel cereblon E3 ligase modulator, targets acute myeloid leukemia blasts and leukemia stem cells. *Blood*. 2021;137:661–677.
173. Matyskiela ME, Lu G, Ito T, et al. A novel cereblon modulator recruits GSPT1 to the CRL4CRBN ubiquitin ligase. *Nature*. 2016;535:252–257.
174. Zeng AGX, Bansal S, Jin L, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med*. 2022;28:1212–1223.
175. Diaz Rohena D, Slawin B, Ravikrishnan J, et al. Targeting venetoclax resistant CLL using a protac-based BCL-2/BCL-XL degrader. *Blood*. 2022;140(Supplement 1):497–498.
176. Wang J, Yu X, Gong W, et al. EZH2 noncanonically binds cMyc and p300 through a cryptic transactivation domain to mediate gene activation and promote oncogenesis. *Nat Cell Biol*. 2022;24:384–399.
177. Basheer F, Giotopoulos G, Meduri E, et al. Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *J Exp Med*. 2019;216:966–981.
178. Sung PJ, Sidoli S, Riedel SS, et al. FLT3 inhibition downregulates EZH2 in AML and promotes myeloid differentiation. *Blood*. 2021;138(Supplement 1):785–785.
179. Tanaka S, Miyagi S, Sashida G, et al. Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood*. 2012;120:1107–1117.
180. Nam AS, Chaligne R, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nature Rev Genet*. 2021;22:3–18.
181. Turkalj S, Jakobsen NA, Groom A, et al. GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells. *Cell Stem Cell*. 2023;30:722–740.
182. Goardon N, Marchi E, Atzberger A, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*. 2011;19:138–152.
183. Quek L, Otto GW, Garnett C, et al. Genetically distinct leukemic stem cells in human CD34– acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *J Exp Med*. 2016;213:1513–1535.
184. Jakobsen NA, Turkalj S, Stoilova B, et al. Single-cell analysis of human clonal hematopoiesis identifies distinct impact of DNMT3A and TET2 mutations on hematopoietic differentiation. *Blood*. 2022;140(Supplement 1):2227–2228.
185. Nakagawa MM, Inagaki R, Kuroda Y, et al. Distinct pathogenesis of clonal hematopoiesis revealed by single cell RNA sequencing integrated with highly sensitive genotyping method. *Blood*. 2022;140(Supplement 1):2881–2882.
186. Nam AS, Dusaj N, Izzo F, et al. Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. *Nat Genet*. 2022;54:1514–1526.

A.2 Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging

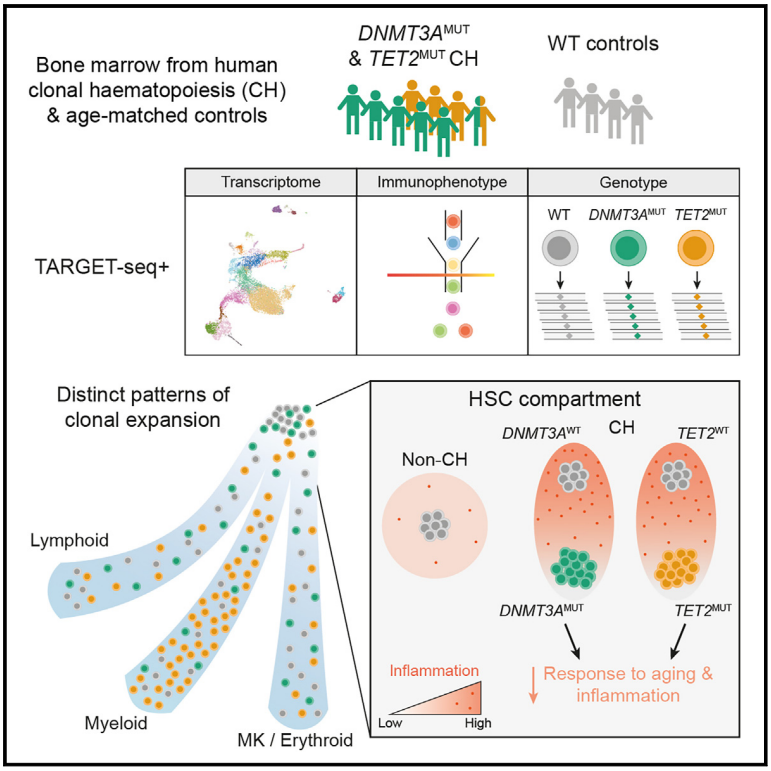
pdfpages

Cell Stem Cell

Article

Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging

Graphical abstract



Highlights

- High-fidelity single-cell genotyping and transcriptomes of human clonal hematopoiesis
- Most clonal expansion occurs in HSCs; *TET2*^{MUT} clones also have a myeloid bias
- HSCs from CH samples show increased inflammation and aging-related gene expression
- These deleterious effects are attenuated in *DNMT3A*^{MUT} and *TET2*^{MUT} HSCs

Authors

Niels Asger Jakobsen, Sven Turkalj, Andy G.X. Zeng, ..., John E. Dick, Stephanie Z. Xie, Paresh Vyas

Correspondence

paresh.vyas@imm.ox.ac.uk

In brief

Vyas and colleagues show *DNMT3A* and *TET2* mutation-specific patterns of clonal expansion in human clonal hematopoiesis (CH). Hematopoietic stem cells (HSCs) in CH bone marrow have heightened transcriptional responses to inflammation. However, these responses are attenuated in CH mutant HSCs, and this could contribute to their clonal advantage.

Jakobsen et al., 2024, Cell Stem Cell *31*, 1–18
August 1, 2024 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.stem.2024.05.010

CellPress

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Cell Stem Cell

CellPress
OPEN ACCESS

Article

Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging

Niels Asger Jakobsen,[1,2,10] Sven Turkalj,[1,2,10] Andy G.X. Zeng,[3,4,10] Bilyana Stoilova,[1] Marlen Metzner,[1] Susann Rahmig,[1] Murtaza S. Nagree,[3] Sayyam Shah,[3] Rachel Moore,[1] Batchimeg Usukhbayar,[1] Mirian Angulo Salazar,[1] Grigore-Aristide Gafencu,[1] Alison Kennedy,[1,5] Simon Newman,[6,7] Benjamin J.L. Kendrick,[6,7] Adrian H. Taylor,[6,7] Rasheed Afinowi-Luitz,[6,7] Roger Gundle,[6,7] Bridget Watkins,[6] Kim Wheway,[6] Debra Beazley,[6] Alex Murison,[3] Alicia G. Aguilar-Navarro,[8] Eugenia Flores-Figueroa,[8] Stephanie G. Dakin,[6] Andrew J. Carr,[6,7] Claus Nerlov,[1] John E. Dick,[3,4,11] Stephanie Z. Xie,[3,11] and Paresh Vyas[1,2,9,11,12,*]

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[2]Oxford Centre for Haematology, NIHR Oxford Biomedical Research Centre, Oxford, UK
[3]Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
[4]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[5]Wellcome - MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
[6]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, Oxford, UK
[7]Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[8]Unidad de Investigación Médica en Enfermedades Oncológicas, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico
[9]Department of Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[10]These authors contributed equally
[11]Senior author
[12]Lead contact
*Correspondence: paresh.vyas@imm.ox.ac.uk
https://doi.org/10.1016/j.stem.2024.05.010

SUMMARY

Clonal hematopoiesis (CH) arises when hematopoietic stem cells (HSCs) acquire mutations, most frequently in the *DNMT3A* and *TET2* genes, conferring a competitive advantage through mechanisms that remain unclear. To gain insight into how CH mutations enable gradual clonal expansion, we used single-cell multi-omics with high-fidelity genotyping on human CH bone marrow (BM) samples. Most of the selective advantage of mutant cells occurs within HSCs. *DNMT3A*- and *TET2*-mutant clones expand further in early progenitors, while *TET2* mutations accelerate myeloid maturation in a dose-dependent manner. Unexpectedly, both mutant and non-mutant HSCs from CH samples are enriched for inflammatory and aging transcriptomic signatures, compared with HSCs from non-CH samples, revealing a non-cell-autonomous effect. However, *DNMT3A*- and *TET2*-mutant HSCs have an attenuated inflammatory response relative to wild-type HSCs within the same sample. Our data support a model whereby CH clones are gradually selected because they are resistant to the deleterious impact of inflammation and aging.

INTRODUCTION

Somatic mutations acquired in long-lived stem cells may provide a substrate for clonal selection, resulting in somatic mosaicism.[1,2] With aging, somatic mosaicism becomes common in phenotypically normal human tissues,[3–12] including the blood, where it is termed clonal hematopoiesis (CH).[13–16] CH is associated with an elevated risk of myeloid malignancy, cardiovascular disease, and other adverse outcomes,[14–18] and these risks increase with clone size.[17–22] Though multiple mechanisms may contribute to clonal expansion,[23] the biological principles remain unclear.

Interestingly, ~70% of CH cases are associated with mutations in *DNMT3A* and *TET2*.[14–17,19,24] DNMT3A, a *de novo* DNA methyltransferase, catalyzes the conversion of cytosine to 5-methylcytosine (5mC), usually in CpG dinucleotides.[25] TET2 is a dioxygenase that catalyzes the conversion of 5mC to 5-hydroxymethylcytosine (5hmC) and other oxidized derivatives.[26,27] This reaction is the first step in DNA demethylation, although 5hmC can also play a regulatory role.[28,29] Both proteins have been associated with additional functions beyond DNA modification.[30–37] In CH, *DNMT3A* mutations are predominantly heterozygous, scattered throughout the three functional

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

domains, and predicted to cause loss of function (LoF). By contrast, ~60% of *DNMT3A* mutations in acute myeloid leukemia affect the R882 residue in the methyltransferase domain. *TET2* mutations are missense or truncating variants distributed across the coding region and are predicted to inhibit or abolish the enzyme's catalytic activity.

In mice, *Dnmt3a*$^{-/-}$ hematopoietic stem cells (HSCs) outcompete wild-type (WT) HSCs in competitive transplants, which has been attributed to increased RNA expression of multipotency and self-renewal genes.[38] Similarly, *Tet2*$^{-/-}$ and *Tet2*$^{+/-}$ HSCs have a competitive advantage in transplantation assays.[39–43] HSC differentiation is also altered, where deletion of *Dnmt3a* increases megakaryocyte-erythroid (MEP) and decreases myelomonocytic cells,[44] whereas *Tet2* loss confers a myeloid bias[41] with reduction of MEP and lymphoid progenitors.[44] Notably, infection and inflammation increase clonal advantage of *Dnmt3a*$^{-/-}$, *Tet2*$^{-/-}$, and *Tet2*$^{+/-}$ HSCs, which is abrogated by deletion of several cytokine receptors.[45–50]

However, murine studies may not capture the complexities of human CH, which arises when a single cell acquires a mutation conferring selective advantage, leading to gradual clonal expansion over time. Indeed, recent studies using population genetic modeling, single-cell phylogenetic analysis, and longitudinal sampling have estimated that *DNMT3A* and *TET2* mutant CH clones expand by about 5%–20% per year and are acquired decades before reaching a substantial clone size.[51–53] By contrast, mice are kept in controlled environments, have short lifespans, and murine studies often assay hematopoiesis after transplantation or introduction of a mutant allele in all blood cells.

In humans, detailed single-cell analyses have been performed in *DNMT3A*^{R882}-mutated CH from patients treated for myeloma,[54] and in cord blood hematopoietic stem and progenitor cells (HSPCs) where *TET2* was experimentally deleted.[55,56] In both settings, differentiation defects were observed, with engineered *TET2*$^{-/-}$ cells having a competitive advantage *in vivo* in immunodeficient mice. However, it remains unclear how *DNMT3A* and *TET2* mutations confer a gradual clonal advantage in age-associated CH. To address this question, we implemented an optimized single-cell multi-omics method combining high-fidelity genotyping with transcriptional profiling to separately analyze mutant and non-mutant HSPCs in humans with CH.

RESULTS

BM sampling from individuals with age-related CH without prior hematological malignancy

To obtain bone marrow (BM) samples from individuals with CH without the confounding effects of co-existing malignancy, we collected samples from 195 individuals with normal blood counts undergoing elective total hip replacement surgery for osteoarthritis. To study steady-state CH, we excluded subjects with prior or current hematological cancer, inflammatory arthritis, or systemic steroid use (Table S1). All samples underwent targeted re-sequencing of BM mononuclear cell (MNC) DNA, using a 97-gene panel with a mean sequencing depth of 822× (Figures 1A and S1A; Table S2). 57 individuals (29.2%) had CH with somatic driver mutation(s) at ≥0.02 variant allele frequency (VAF), and an additional 28 individuals (14.3%) had mutation(s) at 0.01–0.02 VAF (Figures 1B and 1C; Table S2). Consistent with prior studies, 69% of CH cases had mutations in *DNMT3A* and *TET2* (Figures 1C and S1B). The median VAF detected was 0.022 (Figure 1D). Most cases had a single mutation, but the frequency of secondary mutations increased with age (Figure 1E).

CH prevalence increased with age (Figures 1F and S1C). Comparison with previous studies[19,57,58] showed the prevalence in our cohort was similar to a study of peripheral blood (PB) DNA from 3,359 individuals in the general population[58] and to another hip replacement surgery cohort[57] (Figures 1F and S1C).

Most studies of CH to date have performed sequencing on PB DNA. To determine whether the sensitivity for CH detection differs between BM and PB, we compared mutation detection in paired PB granulocytic and BM MNC DNA on 72 samples with CH and 27 samples without CH. PB sequencing identified 14 additional mutations not called in BM sequencing. Although VAFs for some mutations differed considerably between BM and PB, there was no significant difference overall (Figure 1G; $p = 0.11$; Wilcoxon signed-rank test) or specifically for *DNMT3A* and *TET2* mutations (Figure 1H). There was no significant difference in the number mutations with ≥0.02 VAF or individuals with CH detected between BM and PB (Figures 1I and 1J). Our analysis suggests somatic mutations are detected with comparable sensitivity in PB and BM and that the frequency of CH is similar between patients undergoing hip replacement surgery and the general population.

HSPC differentiation landscape in *DNMT3A*- and *TET2*-mutant CH

As mutant cells in CH occur at low frequency, a high-fidelity method for distinguishing mutant and WT cells within the same sample is needed to accurately study the consequences of CH mutations. To address this, we optimized TARGET-seq,[59] which combines high-fidelity single-cell genotyping with transcriptome sequencing on flow cytometry index-sorted cells. Our new method, TARGET-seq+, incorporates elements of Smart-seq3 chemistry[60] to increase transcript detection sensitivity. We compared TARGET-seq and TARGET-seq+ on JURKAT cells and primary human lineage$^-$ (Lin$^-$) CD34$^+$ HSPCs (Figure S1D). Sequencing metrics were comparable (Figure S1E), but TARGET-seq+ yielded a higher proportion of cells passing quality filters (Figure S1F) and increased the number of genes detected per cell by 13.5% in JURKAT cells and by 19.0% in HSPCs (Figures S1G and S1H). Increased detection was observed in both frequently and lowly expressed genes (Figure S1I). Consistent with better transcript capture, cell-to-cell correlations of transcript levels improved with TARGET-seq+ (Figure S1J).

We applied TARGET-seq+ to 9 CH samples and 4 age-matched controls without known driver mutations (non-CH samples) (Figure 2A; Table S3). We focused on *DNMT3A*-mutant (*DNMT3A*^{MUT}) and *TET2*-mutant (*TET2*^{MUT}) CH by selecting 5 cases with heterozygous *DNMT3A* LoF mutations, 3 with *TET2* LoF mutations, and one case with mutations in both *DNMT3A* and *TET2*. VAFs in BM ranged from 0.061 to 0.366 (Table S3). In two CH cases, there were additional mutations in other genes at VAF < 0.02 (Table S3). We sorted a mean of 1,071 cells/sample (range 348–1,824), composed of purified Lin$^-$CD34$^+$ HSPCs, further enriched for primitive Lin$^-$CD34$^+$CD38$^-$ HSCs and

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

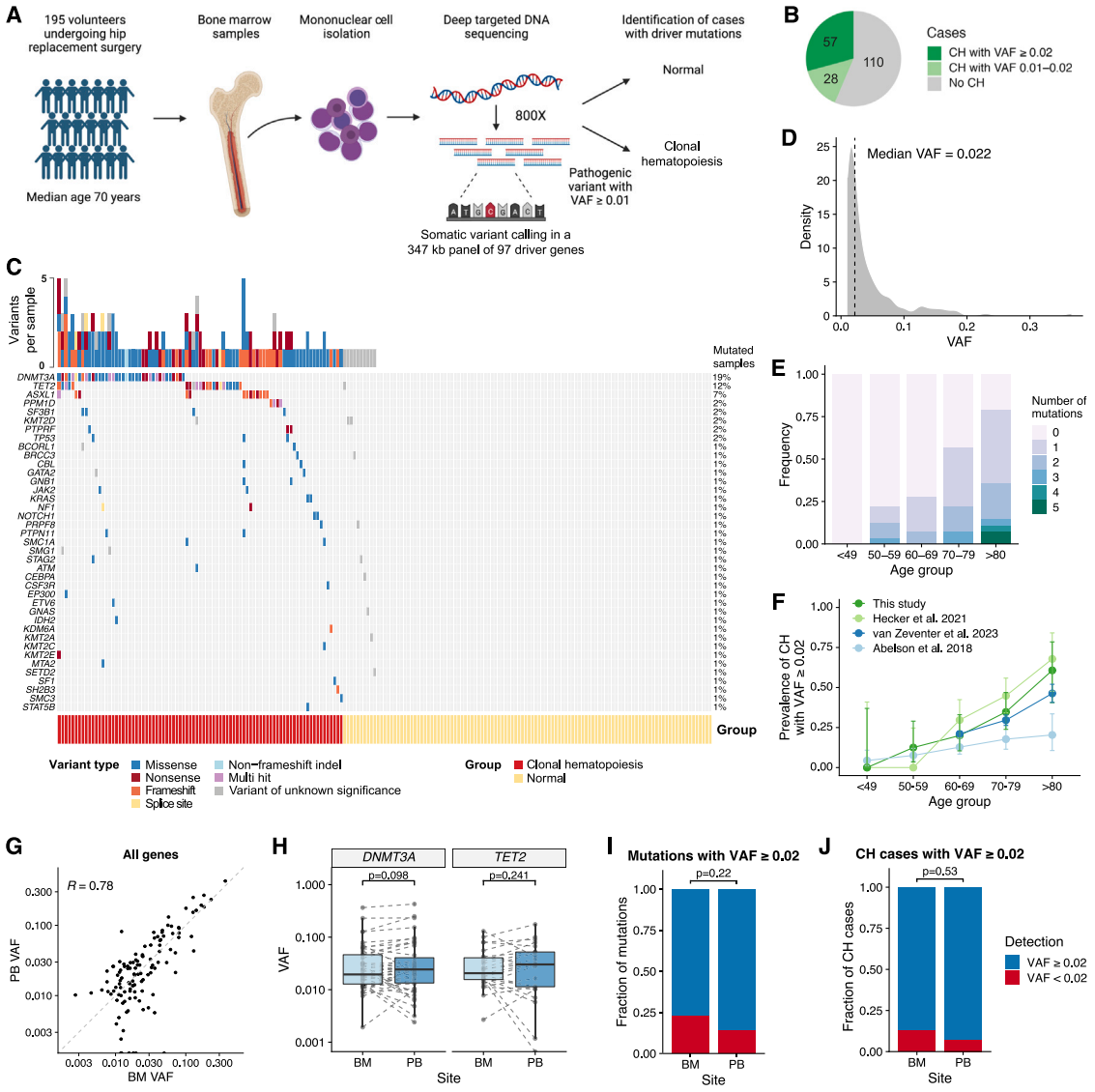


Figure 1. Identification of age-related clonal hematopoiesis in individuals undergoing hip replacement surgery

(A) Experimental design for identifying individuals with clonal hematopoiesis (CH).

(B) Fraction of samples in the cohort with driver mutation(s) at 0.01–0.02 or ≥0.02 VAF.

(C) Landscape of somatic variants observed in the cohort. Each row represents a gene, and each column represents a study participant. Top bar plot indicates the number of mutations per sample. Variants are classified as pathogenic or variants of unknown significance (VUSs) (STAR Methods). Samples with ≥1 pathogenic driver mutation were categorized as having CH.

(D) Distribution of VAFs in all mutations observed across the cohort.

(E) Frequency of mutations detected per individual by age group.

(F) Prevalence of CH with driver mutation(s) ≥0.02 VAF by age. BM sequencing data from this study are compared with another hip replacement cohort (Hecker et al.[57]; n = 109 BM and n = 91 PB samples; green colors) and with two studies of the general population[19,58] (blue colors). Error bars represent 95% confidence intervals (CIs).

(G) Comparison of VAFs for 128 mutations in paired BM and PB samples. Mutations detected with VAF ≥0.01 in either sample type were included. The dashed line shows the line of equality where BM VAF is equal to PB VAF. R indicates the Pearson correlation coefficient.

(H) Pairwise comparison of VAFs for mutations in *DNMT3A* (n = 35) and *TET2* (n = 19). Significance calculated by Wilcoxon signed-rank test.

(I and J) Proportion of mutations (I) or CH cases (J) detected with ≥0.02 VAF in BM or PB DNA (n = 83 cases with paired BM and PB data). Significance calculated by Fisher's test.

multipotent progenitors (MPPs), combined with CD34$^-$ mature cells (Figure S1K). Approximately 40%–45% of cells across all samples were Lin$^-$CD34$^+$CD38$^-$, a similar percentage were Lin$^-$CD34$^+$CD38$^+$, and ~10% were CD34$^-$ (Figure S1L). 95.0% of sorted cells (13,247/13,939) were included for further analysis after quality filtering (Figure S1M). We detected ~10^6

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

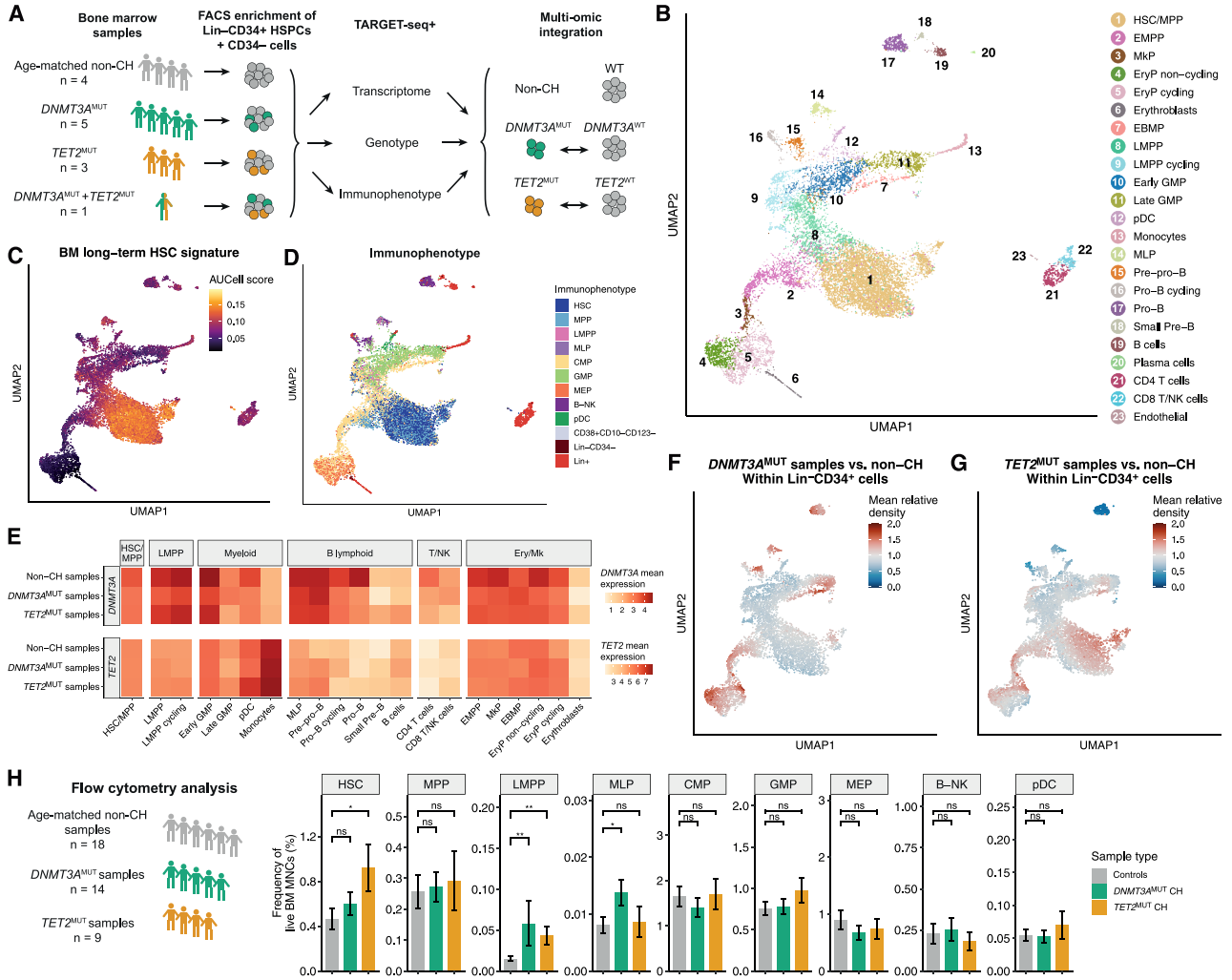


Figure 2. Hematopoietic differentiation trajectory in *DNMT3A* and *TET2*-mutant clonal hematopoiesis

(A) Experimental design for TARGET-seq+ analysis of BM samples from donors with CH and age-matched samples without CH. See Figure S1.

(B) Uniform manifold approximation and projection (UMAP) of integrated single-cell transcriptome data (n = 13,247 cells from 13 donors) colored by cluster annotation. See Figure S2.

(C) UMAP with AUCell enrichment scores for the BM long-term HSC signature.[61]

(D) UMAP colored by cell immunophenotype determined from flow cytometry indexing.

(E) Heatmap of mean $\log_2$(normalized counts) for *DNMT3A* and *TET2* in control, *DNMT3A*-, and *TET2*-mutant samples across hematopoietic cell types.

(F and G) UMAPs colored by the mean density of Lin$^-$CD34$^+$ cells in *DNMT3A*-mutant (F) and *TET2*-mutant (G) CH samples relative to non-CH samples. Relative density >1 indicates that the probability of observing a given cell is greater in CH samples than in non-CH samples.

(H) Flow cytometry analysis on BM samples from non-CH samples and CH samples with either *DNMT3A* or *TET2* mutations present in the largest clone. Data are represented as mean ± SEM. p values calculated by Wilcoxon rank-sum test with Holm-Bonferroni multiple testing correction. *$p < 0.05$, **$p < 0.01$.

GMP, granulocyte-monocyte progenitor; pDC, plasmacytoid dendritic cell progenitor; MkP, megakaryocytic progenitor; EryP, erythroid progenitor; EBMP, eosinophil/basophil/mast cell progenitor; MLP, multi-lymphoid progenitor; B-NK, B/NK cell progenitor.

RNA counts (Figure S1N) and a median of 6,484 genes per cell (Figure S1O). Importantly, these metrics were consistent across samples.

To generate a hematopoietic landscape based on single-cell RNA profiles, we integrated gene expression data across the 13 samples (Figure S2A). We annotated 23 cell clusters using published gene signatures[61–69] and marker genes (Figures 2B, 2C, S2B, and S2C). Immunophenotypic HSPC populations, defined from flow cytometry index data, were often present in multiple transcriptionally defined populations (Figures 2D and S2D), consistent with transcriptional clusters providing a more granular view of hematopoietic cell states.[62,67]

Downstream of HSC/MPPs, a continuum of cell states was observed, with initial separation into lymphoid-primed MPPs (LMPPs) and erythroid/megakaryocytic-primed MPPs (EMPPs) (Figure 2B). A series of cell states with neutrophilic/monocytic and lymphoid potential extended downstream of LMPPs, whereas progressing from the EMPP were erythroid cells and megakaryocytes, consistent with prior data.[64,70] *DNMT3A* and *TET2* were expressed in HSC/MPP, LMPP, and EMPP but

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

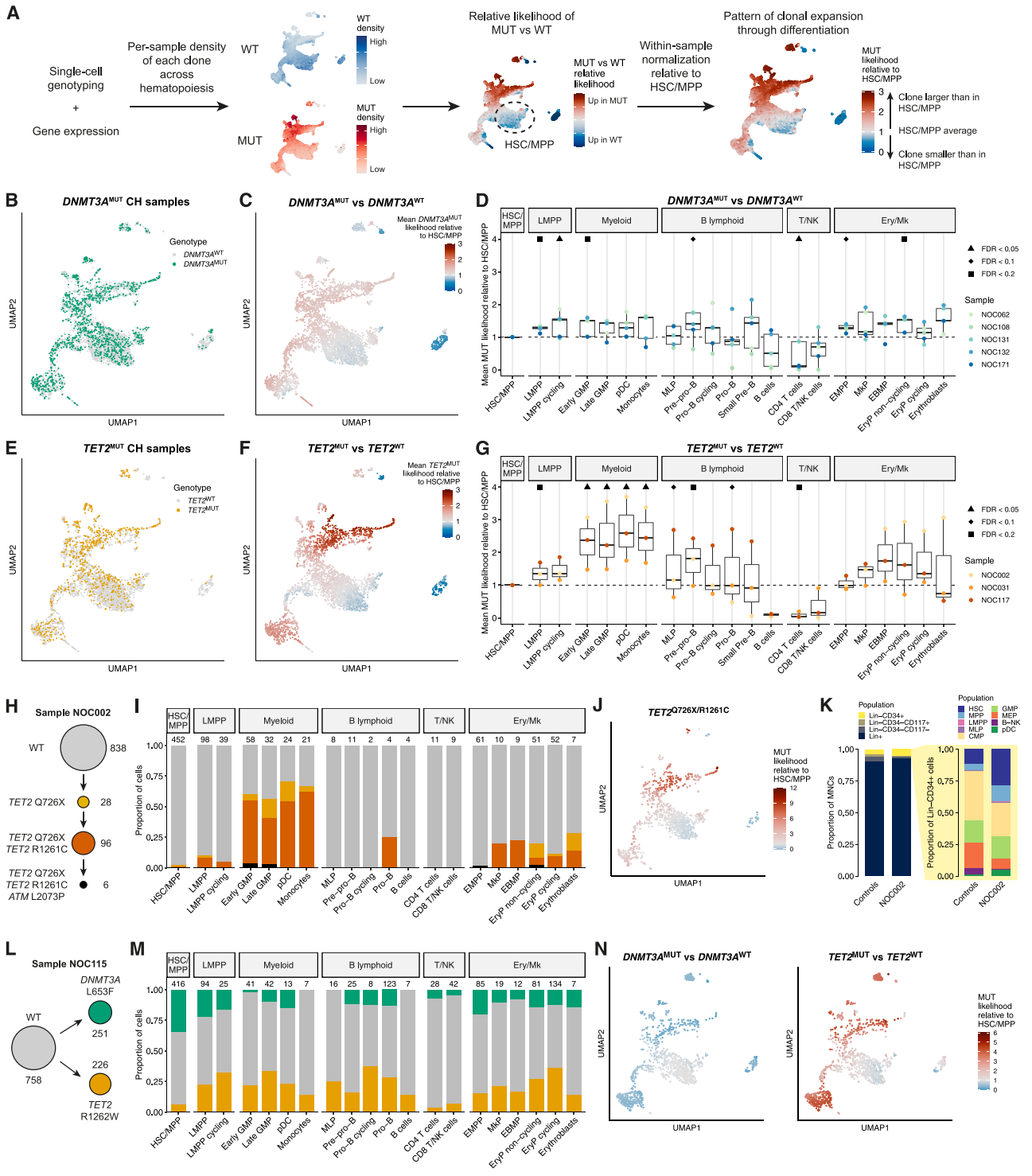


Figure 3. Distinct patterns of clonal expansion of *DNMT3A*- and *TET2*-mutant clones

(A) Strategy for quantifying CH^{MUT} clonal expansion across hematopoietic differentiation. For each sample, MELD and scCODA were used to estimate the density of cells from each clone across the transcriptomic landscape (see Figures S3 and S4 and STAR Methods). A mutant relative likelihood >1 indicates that the probability of a cell being mutant is greater than in the HSC/MPP, whereas a relative likelihood <1 indicates that the probability is lower than in the HSC/MPP.

(B) UMAP of cells from *DNMT3A*^{MUT} CH samples (n = 5 samples) colored by genotype.

(C) UMAP of cells from *DNMT3A*^{MUT} CH samples colored by the mean likelihood of cells being *DNMT3A*^{MUT} relative to the average within HSC/MPP. The mean value across 5 samples is shown.

(legend continued on next page)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

decreased in late lymphoid and erythroid lineages (Figure 2E). Notably, *TET2* expression increased during myeloid maturation, in contrast to *DNMT3A*.

We first asked whether the size of HSPC compartments is perturbed in CH, given that their frequencies change with age[71] and in myeloid disease.[72–75] Comparing the frequency of cells across the transcriptional landscape between CH and non-CH samples (STAR Methods) showed moderate differences in various populations (Figures 2F and 2G), but these were not significant (at false discovery rate [FDR] < 0.2 using scCODA[76]; Figures S2E and S2F). Interestingly, conventional immunophenotyping on a larger sample set showed $Lin^-CD34^+CD38^-CD45RA^-CD90^+$ HSCs and $CD49f^+$ long-term HSCs (LT-HSCs) were 2- and 2.2-fold expanded, respectively, in *TET2*-mutant relative to non-CH samples (Figures 2H and S2G). Additionally, *DNMT3A*- and *TET2*-mutant CH samples showed a 3.9- and 2.9-fold increase in the rare $Lin^-CD34^+CD38^-CD45RA^+$ LMPP population, respectively (Figure 2H), which constitutes a minor fraction of the transcriptionally defined LMPP and multi-lymphoid progenitors (MLPs) (Figure S2D). Overall, the size of HSPC compartments is moderately perturbed in *DNMT3A*- and *TET2*-mutant CH.

Distinct patterns of clonal expansion of *DNMT3A*- and *TET2*-mutant clones

Next, we integrated genotypes with transcriptional profiles to identify mutant and WT cells within the same sample. To accurately call genotypes while accounting for noise in PCR amplification and sequencing, we simultaneously genotyped cells from a WT control sample, which allowed us to determine the error rate at each locus (Figures S3A and S3B). Overall, *DNMT3A* and *TET2* loci were successfully amplified in 97.7% and 98.1% of cells, respectively, resulting in clonal assignment for 93.0% of CH cells (Figure S3C). Genotyping rates for *DNMT3A* and *TET2* were high across all samples (Figure S3C) and cell types (Figure S3D).

Loss of one allele during PCR amplification, termed allelic dropout (ADO), is common in single-cell genotyping and can compromise the accuracy of downstream analysis due to misassignment of heterozygous mutant cells as WT. We used two strategies to estimate ADO. In four loci, heterozygous germline single nucleotide polymorphisms (SNPs) showed ADO rates of 5.3%–13.8% (Figures S3E–S3H). For the remaining loci, where we inferred ADO rates from the frequency of cells that appeared homozygous mutant (i.e., where we assumed there was ADO of the WT allele; Figure S3I), these rates were 0%–13.5% (Figure S3J). The frequency of mutant cells detected by TARGET-seq+ correlated well with estimates from bulk analysis of Lin^-CD34^+ cells ($R = 0.98$), validating our single-cell genotyping (Figure S3K). A potential alternative way to assess genotyping accuracy would be to quantify expression of a gene specific for a certain genotype. It has been suggested that *TCL1A* is upregulated in $TET2^{MUT}$ HSCs among individuals without the rs2887399 variant.[77] Across $TET2^{MUT}$ CH samples, 0.36%–2.5% of WT HSC/MPPs had >2 reads mapping the *TCL1A* gene ($TCL1A^{hi}$), a higher rate compared with non-CH samples (Figure S3L). However, these $TCL1A^{hi}TET2^{WT}$ cells transcriptionally resembled other $TET2^{WT}$ cells rather than $TET2^{MUT}$ cells (Figures S3M–S3O). Therefore, in our dataset, $TCL1A^{hi}$ status was not specific to $TET2^{MUT}$ HSC/MPPs.

In all subjects, *DNMT3A* or *TET2* mutations were present in the founder clone, allowing us to determine the effect of mutations in a WT background (Figures S4A–S4H). To reveal how mutant clones expand or contract with differentiation, we projected genotypes onto the transcriptomic landscape and compared the density of mutant and WT cells (referred to as the mutant clone likelihood) to quantify how clone size changed relative to HSC/MPP (Figure 3A; STAR Methods).

$DNMT3A^{MUT}$ cells were intermingled with $DNMT3A^{WT}$ cells, both within individual samples (Figures S4A–S4E) and in the integrated dataset (Figure 3B), indicating that they shared a similar differentiation trajectory and that mutant cells did not create novel transcriptional states. In HSC/MPPs, the $DNMT3A^{MUT}$ clone size was highly variable, ranging from 3.4% to 73.3% (Figures S4A–S4E). Changes in $DNMT3A^{MUT}$ clone size through differentiation were modest (Figures 3C and 3D). Relative to HSC/MPPs, the mean mutant clone size was approximately 30% larger in early EMPP and LMPP populations (Figures 3C, 3D, and S4I). Clone sizes were largely maintained at later stages of differentiation, except in T cells, where clones were on average 50% smaller than in HSC/MPPs. Aside from this, there was no consistent lineage bias, but there was notable variability among individuals. While one sample showed almost complete absence of $DNMT3A^{MUT}$ cells in lymphoid cells (Figure S4A), consistent with previous data from PB populations,[78–80] another showed depletion of $DNMT3A^{MUT}$ cells in myeloid progenitors

(D) Mean $DNMT3A^{MUT}$ clone likelihood in each cluster relative to the HSC/MPP, computed using MELD. Each dot represents a $DNMT3A^{MUT}$ CH sample. Boxplots display the median and interquartile range. Symbols above indicate whether a significant difference in clone size relative to the HSC/MPP was detected using scCODA.
(E) UMAP of cells from $TET2^{MUT}$ CH samples ($n = 3$ samples) colored by genotype.
(F) UMAP of cells from $TET2^{MUT}$ CH samples colored by the mean likelihood of cells being $TET2^{MUT}$ relative to the average within HSC/MPP. The mean value across 3 samples is shown.
(G) Same as in (D), but for the $TET2^{MUT}$ clone likelihoods across $TET2^{MUT}$ CH samples.
(H) Clonal structure for the NOC002 sample. Cell numbers in each clone are indicated.
(I) Clonal composition within each cluster for sample NOC002. Each clone is colored as in (H). The number of cells analyzed in each cluster is shown above.
(J) UMAP showing the likelihood of cells being in the double $TET2^{MUT}$ clone ($TET2^{Q726X/R1261C}$) relative to the average within HSC/MPP in sample NOC002.
(K) Immunophenotypic BM compartment sizes comparing sample NOC002 with the median from 18 age-matched control samples. Left-hand bars: compartments as a proportion of total BM MNCs; right-hand bars: HSPC compartments within Lin^-CD34^+ cells.
(L) Clonal structure for the NOC115 sample. The *DNMT3A* and *TET2* mutations were mutually exclusive in single-cell genotyping. Cell numbers in each clone are indicated.
(M) As in (I) but for sample NOC115. Each clone is colored as in (L).
(N) UMAPs showing the likelihood of cells being in the $DNMT3A^{MUT}$ (left) and $TET2^{MUT}$ (right) clones relative to the average within HSC/MPP in sample NOC115.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

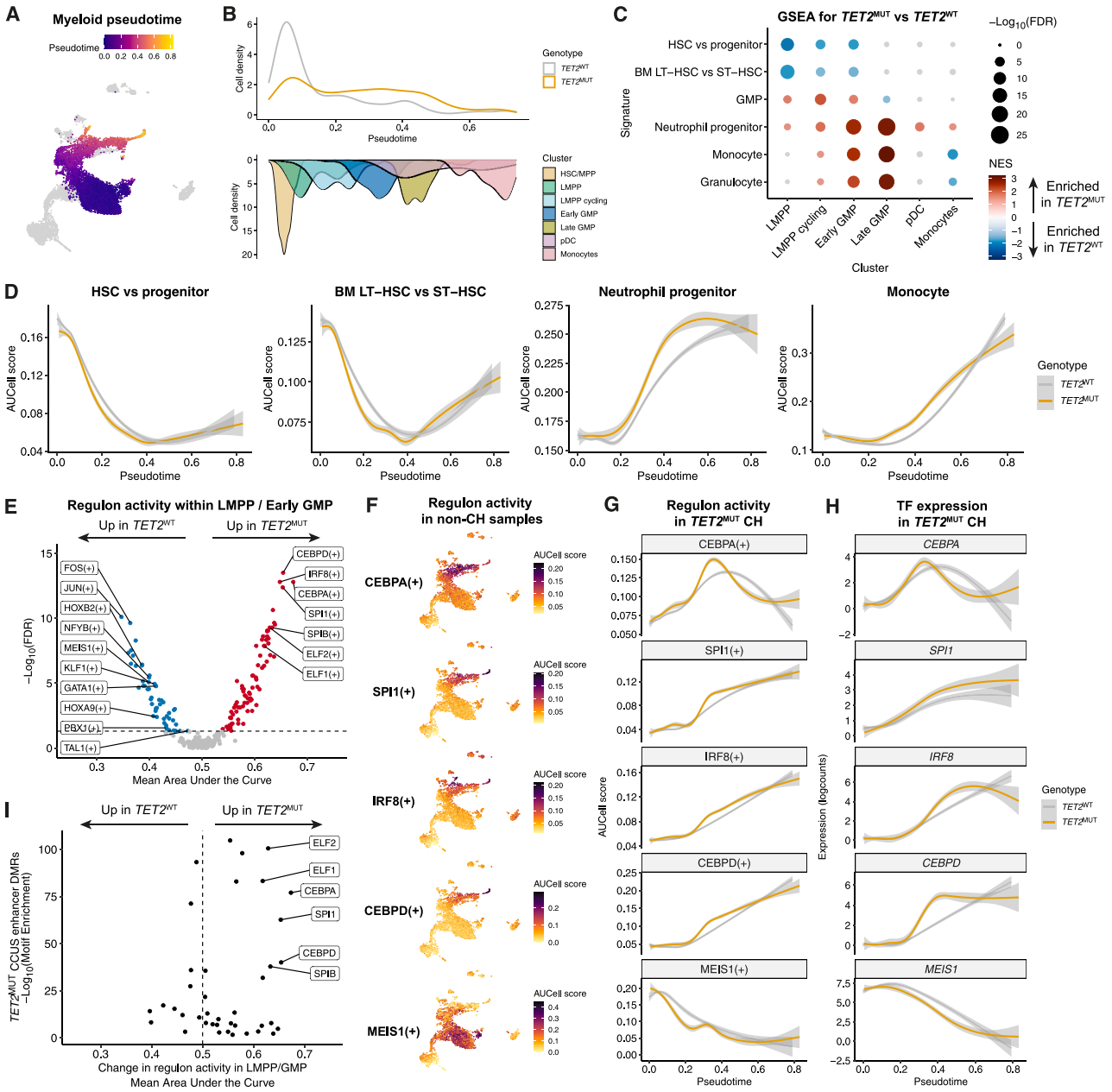


Figure 4. *TET2*-mutant clones lead to dysregulated myeloid differentiation

(A) UMAP showing the myeloid differentiation trajectory with cells colored by pseudotime score.

(B) Top: density plot showing the distribution of *TET2*^{WT} and *TET2*^{MUT} cells through pseudotime in the myeloid lineage. Cells sorted from the total Lin^{-}CD34^{+} fluorescence-activated cell sorting (FACS) gate were downsampled to an equal number of cells per sample (n = 178 cells from each of the 4 samples). Bottom: density of cells in each cluster along pseudotime.

(C) GSEA against hematopoietic signatures comparing *TET2*^{MUT} versus *TET2*^{WT} cells (n = 4 *TET2*^{MUT} CH samples) within myeloid lineage clusters. Differential expression analysis was performed accounting for sample and batch effects. Signatures with FDR > 0.05 are colored gray. Positive normalized enrichment scores (NESs) indicate enrichment in mutant cells. LT-HSC, long-term HSC; ST-HSC, short-term HSC.

(D) Local regression of AUCell expression scores for HSC and myeloid gene signatures along myeloid pseudotime, comparing *TET2*^{WT} and *TET2*^{MUT} cells. Shading indicates the 95% CI.

(E) Volcano plot showing differentially expressed regulons between *TET2*^{MUT} and *TET2*^{WT} cells within the LMPP cycling and early GMP clusters in *TET2*^{MUT} CH samples. FDR-corrected p values calculated by linear mixed model test accounting for sample effects. x axis shows the mean of the change in regulon activity (area under the receiver operator curve) calculated within each sample.

(F) UMAPs showing activity of the indicated regulons across the hematopoietic landscape within non-CH samples.

(legend continued on next page)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

(Figure S4D). Overall, *DNMT3A*^{MUT} clonal expansion occurred primarily in HSCs and early MPPs without consistent lineage bias.

TET2^{MUT} cells also intermingled with *TET2*^{WT} cells (Figure 3E). Clone size within HSC/MPPs was 1.1%–32.8% (Figures S4F–S4H). In contrast to *DNMT3A*^{MUT} clones, there was pronounced expansion of *TET2*^{MUT} clones downstream of HSC/MPPs during myelopoiesis (Figures 3F, 3G, and S4I). Indeed, *TET2*^{MUT} clones were on average 2.5- to 3-fold larger in granulocyte-monocyte progenitors (GMPs) compared with HSC/MPPs. In 2 of 3 samples, the mutant clone also expanded within erythroid progenitors (Figures S4G and S4H). There was heterogeneity in the contribution of *TET2*^{MUT} cells to lymphoid progenitors, where mutant cells were depleted in 2 cases (Figures S4F and S4H), but the *TET2*^{MUT} clone constituted 90% of B cell progenitors in another individual (Figure S4G). By contrast, *TET2*^{MUT} cells were almost absent from mature B and T cells in all samples, suggesting inability to complete lymphoid differentiation.

Next, to understand the differentiation potential of heterozygous versus homozygous *TET2*^{MUT} clones, we studied one sample where two *TET2* mutations were acquired sequentially in a linear clonal structure (Figure 3H). The single- and double-*TET2*^{MUT} clones each contributed to only 1.1% of HSC/MPPs. While the single-*TET2*^{MUT} clone expanded 3- to 4-fold in erythroid and myeloid progenitors, the double-*TET2*^{MUT} clone dramatically outcompeted the single mutant clone during myelopoiesis, contributing to >50% of GMPs, plasmacytoid dendritic cells (pDCs), and monocytes (Figures 3I and 3J), suggesting a dose-dependent effect of *TET2* LoF on myeloid clonal expansion. These findings were validated by droplet digital PCR (ddPCR) genotyping of immunophenotypic populations (Figure S4J). Interestingly, though 97.6% of HSC/MPPs were *TET2*^{WT}, there was a 2.6-fold expansion of immunophenotypic HSC/MPPs in this sample relative to non-CH samples (Figure 3K). This raises the question of whether *TET2*^{MUT} CH might increase WT HSC/MPP cell numbers in a non-cell-autonomous manner.

Finally, we had the opportunity to study clonal competition between *DNMT3A*^{MUT} and *TET2*^{MUT} cells in an individual with co-existing, independent *DNMT3A*^{MUT} and *TET2*^{MUT} clones (Figure 3L). Interestingly, the *DNMT3A*^{MUT} clone was 5 times larger than the *TET2*^{MUT} clone within HSC/MPPs, but the *TET2*^{MUT} clone became 4 times larger than the *DNMT3A*^{MUT} clone within GMPs (Figures 3M and 3N). Notably, the *TET2*^{MUT} clone was also larger than the *DNMT3A*^{MUT} clone in erythroid and lymphoid progenitors (Figure 3M), observations once again validated by ddPCR (Figure S4K).

In summary, *DNMT3A*^{MUT} and *TET2*^{MUT} clones showed distinct patterns of clonal expansion across differentiation. The selective advantage of *DNMT3A*^{MUT} clones occurs mainly in HSCs and early MPPs, whereas *TET2*^{MUT} clones expand in HSCs and further through differentiation, especially in myelopoiesis.

Transcriptional basis for dysregulated myeloid differentiation of *TET2*-mutant clones

As *TET2*^{MUT} clonal expansion was most pronounced in the myeloid lineage, we further explored the transcriptional basis of this phenotype in individuals with *TET2*^{MUT} CH. We first compared the density distributions of *TET2*^{MUT} and *TET2*^{WT} cells along the differentiation trajectory from HSCs to mature myeloid cells (Figure 4A). *TET2*^{MUT} cells accumulated, particularly at the progenitor stage, within the cycling LMPP and GMP clusters (Figure 4B).

TET2^{MUT} myeloid progenitor expansion might arise due to increased myeloid differentiation from HSCs and/or delayed terminal maturation of mutant myeloid progenitors. To investigate this, we performed gene set enrichment analysis (GSEA) using published HSPC signatures[61,62,66] validated in our dataset (Figure S5A) and examined transcriptional differences between *TET2*^{MUT} and *TET2*^{WT} cells within LMPPs and GMPs. *TET2*^{MUT} progenitors were negatively enriched for HSC signatures (Figure 4C), suggesting decreased retention of stemness programs. Conversely, *TET2*^{MUT} progenitors, particularly GMPs, were enriched for neutrophil progenitor and mature neutrophil/monocyte signatures (Figure 4C). Importantly, these patterns were consistent across multiple individual samples (Figure S5B). Concordantly, CD38 and CD45RA surface protein expression was higher in *TET2*^{MUT} than *TET2*^{WT} LMPPs (Figure S5C), consistent with our previous observations that higher expression of these markers enriches for myeloid potential in LMPPs.[67] Megakaryocytic-erythroid signatures were negatively enriched, particularly in *TET2*^{MUT} cycling LMPPs (Figure S5D), concordant with their myeloid bias. *TET2*^{MUT} progenitors also upregulated signatures associated with cell cycle (in cycling LMPP), oxidative phosphorylation, cytokine signaling, and innate immune effector function (in GMP) (Figure S5E). Conversely, cell cycle and oxidative phosphorylation signatures were negatively enriched in *TET2*^{MUT} Pro-B/Pre-B cells (Figure S5E), providing a potential explanation for the relative depletion of *TET2*^{MUT} cells in mature B cells (Figure 3G).

To explore these transcriptional alterations throughout the continuum of myeloid differentiation, we compared AUCell scores[82] for the same HSPC signatures between *TET2*^{MUT} and *TET2*^{WT} cells across myeloid pseudotime. HSC-related genes were downregulated earlier in *TET2*^{MUT} cells, while neutrophil progenitor and monocyte signatures showed premature upregulation in *TET2*^{MUT} cells (Figure 4D). Concordantly, exemplar genes expressed in mature myeloid cells, including *MPO*, *NKG7*, *KLF4*, and *RBM47*, showed premature expression in *TET2*^{MUT} progenitors (Figure S5F). This suggests that early *TET2*^{MUT} lympho-myeloid progenitors retain less HSC and non-myeloid programs, committing more rapidly to myelopoiesis.

To identify potential drivers of dysregulated myeloid differentiation in *TET2*^{MUT} progenitors, we used pySCENIC to identify

(G) Local regression of regulon activity through myeloid pseudotime, comparing *TET2*^{MUT} and *TET2*^{WT} cells. Shading indicates the 95% CI.
(H) Fitted gene expression values along myeloid pseudotime for the transcription factors shown in (F) and (G) in *TET2*^{MUT} and *TET2*^{WT} cells. Shading indicates the 95% CI.
(I) Enrichment of TF motifs within differentially methylated enhancer regions (DMRs) that are hypermethylated in monocytes from *TET2*-mutant CCUS patients,[81] plotted against the mean change in regulon activity between *TET2*^{MUT} and *TET2*^{WT} cells within the LMPP cycling and early GMP clusters from (E).
See also Figure S5.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

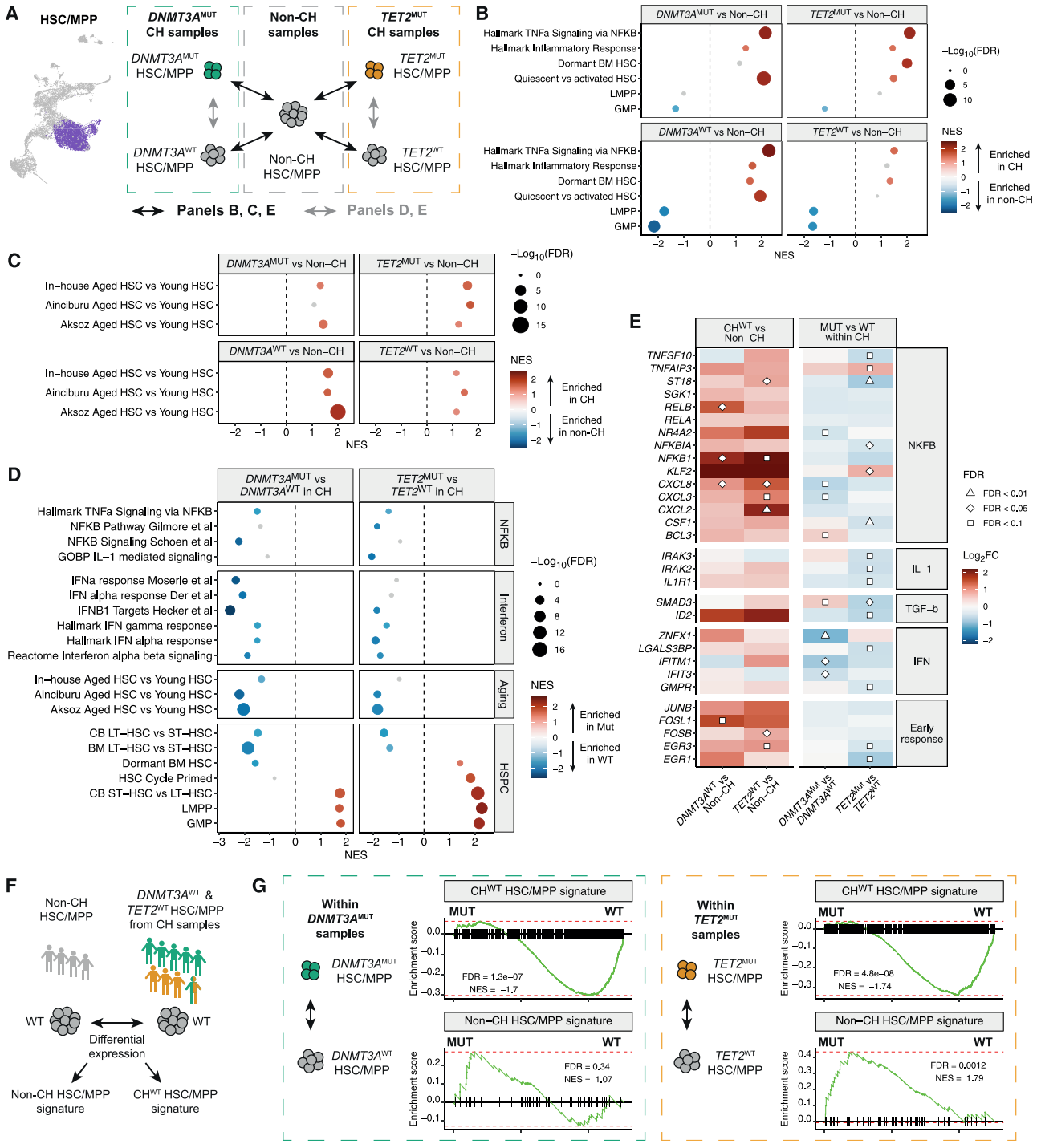


Figure 5. Non-cell-autonomous activation of inflammatory transcriptional programs in clonal hematopoiesis is attenuated in mutant HSCs

(A) Strategy for differential gene expression analysis between HSC/MPPs from CH samples and HSC/MPPs from age-matched non-CH samples (black arrows; B, C, and E), and between CH^{MUT} and CH^{WT} HSC/MPPs within CH samples (gray arrows; D and E).

(B) GSEA against inflammatory and hematopoietic lineage signatures comparing CH^{WT} or CH^{MUT} HSC/MPPs versus non-CH HSC/MPPs. Differential expression analysis was performed accounting for sample, age, and batch effects. Left: 5 *DNMT3A*^{MUT} samples (n = 1,139 *DNMT3A*^{WT} cells, n = 409 *DNMT3A*^{MUT} cells) versus 4 non-CH samples (n = 1,279 cells). Right: 3 *TET2*^{MUT} samples (n = 1,239 *TET2*^{WT} cells, n = 222 *TET2*^{MUT} cells) versus 4 non-CH samples (n = 1,279 cells). The double-mutant NOC115 sample was excluded. Signatures with FDR > 0.2 are colored gray. Positive NES values indicate enrichment in CH samples.

(C) As in (B) but showing GSEA against aged HSC signatures derived from the in-house dataset and from two additional studies comparing aged and young human HSCs.[84,85]

(legend continued on next page)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

transcription factors (TFs) and their putative downstream target genes (i.e., regulons). We compared regulon expression between *TET2*^{MUT} and *TET2*^{WT} cells within LMPP and early GMP clusters, where lymphoid and myeloid lineages diverge. The canonical myeloid TFs CEBPD, CEBPA, IRF8, SPI1, and SPIB were more active in *TET2*^{MUT} progenitors (Figure 4E). Conversely, TFs associated with HSC self-renewal (MEIS1, HOXA9, HOXB2, NFYB, and PBX1) and with megakaryocytic-erythroid differentiation (GATA1, TAL1, and KLF1) were less active in *TET2*^{MUT} cells. Regulon activity (Figures 4F and 4G) and TF expression (Figure 4H) of *CEBPA*, *SPI1*, *IRF8*, and *CEBPD* peaked earlier in *TET2*^{MUT} cells along myeloid pseudotime. This quartet of myeloid TFs is required both in early (CEBPA, SPI1, and IRF8) and later (SPI1, IRF8, and CEBPD) stages of myelopoiesis. By contrast, *MEIS1* and its targets were downregulated earlier in *TET2*^{MUT} cells. Interestingly, binding motifs of CEBPA, CEBPD, SPI1, SPIB, and ELF1/2 were also enriched within differentially methylated enhancers in PB granulocytes from patients with *TET2*^{MUT} clonal cytopenia of undetermined significance (CCUS) (Figure 4I).[81] This suggests a link between altered enhancer methylation and dysregulated myeloid TF activity in *TET2*^{MUT} myeloid cells. Overall, these findings show *TET2*^{MUT} myeloid progenitors are biased toward maturation, with accelerated upregulation of mature transcriptional programs.

Non-cell-autonomous activation of inflammatory transcriptional programs in CH is attenuated in mutant HSCs

Humans have an estimated 50,000–200,000 HSCs.[83] In an individual where 1% of HSCs harbor a CH mutation, this represents a 500- to 2,000-fold expansion from a single initiating mutant HSC. Compared with this, changes in clone size downstream of HSCs are modest, implying that the greatest clonal expansion for both *DNMT3A*^{MUT} and *TET2*^{MUT} (CH^{MUT}) clones occurred within long-lived HSCs. At least two hypotheses could explain the relative clonal advantage of CH^{MUT} HSCs over WT HSCs (CH^{WT} HSCs): either CH^{MUT} HSCs have a cell-autonomous (intrinsic) competitive advantage and/or CH^{WT} HSCs are at a competitive disadvantage due to cell-extrinsic differences in the CH BM microenvironment.

To dissect these hypotheses, we first compared gene expression between HSC/MPPs from CH samples (both CH^{MUT} and CH^{WT}) and age-matched samples without CH (Figure 5A). Compared with non-CH HSC/MPPs, CH^{MUT} HSC/MPPs showed enrichment of tumor necrosis factor alpha (TNF-α) signaling via nuclear factor κB (NF-κB), inflammatory response, and HSC quiescence signatures (Figure 5B, top). Strikingly, a similar pattern was observed in CH^{WT} versus non-CH HSC/MPPs (Figure 5B, bottom). By contrast, gene sets for LMPP/GMP states were negatively enriched in CH^{WT} HSC/MPPs, suggesting decreased priming toward differentiation. These data suggest both mutant and WT HSC/MPPs in CH individuals are impacted by an inflammatory milieu.

Prior data indicate that aging is associated with chronic inflammation, increased NF-κB signaling, quiescence, and functional decline in HSCs.[86–88] To derive human aging-associated transcriptional programs, we performed single-nucleus RNA sequencing (snRNA-seq) on human BM HSPCs and performed differential expression between HSCs from young and aged individuals to define a signature of aged HSCs (Figures S6A and S6B). For further validation, we defined additional aged HSC signatures through re-analysis of two human HSPC single-cell RNA-seq (scRNA-seq) datasets (Figures S6C and S6D; Table S6).[84,85] Notably, these aged HSC signatures were enriched in both CH^{MUT} and CH^{WT} HSC/MPPs compared with age-matched non-CH HSC/MPPs (Figure 5C).

We then asked how *DNMT3A* and *TET2* mutations alter gene expression to provide CH^{MUT} HSC/MPPs a fitness advantage within the CH BM environment. Interestingly, TNF-α signaling, NF-κB pathway, interleukin-1 (IL-1) signaling, and interferon (IFN)-α response signatures were all negatively enriched in CH^{MUT} HSC/MPPs compared with their CH^{WT} counterparts (Figure 5D). Furthermore, aged HSC and LT-HSC signatures were also negatively enriched in CH^{MUT} compared with CH^{WT} HSC/MPPs, particularly in individuals with *DNMT3A*^{MUT} CH. These patterns were shared across multiple individual samples, despite some heterogeneity across individuals (Figure S6E). Specific genes associated with interferon, NF-κB, IL-1, transforming growth factor β (TGF-β) signaling, and early response were more highly expressed in CH^{WT} compared with non-CH HSC/MPPs (Figure 5E, left). Conversely, many of the same genes were expressed at lower levels in CH^{MUT} compared with CH^{WT} HSC/MPPs (Figure 5E, right). Interestingly, *TNFAIP3*, encoding A20, which inhibits NF-κB activation, was upregulated in *TET2*^{MUT} HSC/MPP. By contrast, signatures of short-term HSC (ST-HSC), LMPP, and GMP were enriched in CH^{MUT} compared with CH^{WT} HSC/MPPs (Figure 5D), suggesting an inverse relationship between molecular programs underlying inflammation and aging in contrast to lympho-myeloid differentiation priming. Consistent with reduced quiescence, CH^{MUT} HSC/MPPs were enriched for pathways related to mitosis, cell migration, and signaling, particularly in *TET2*^{MUT} cells (Figure S6F). Furthermore, *TET2*^{MUT} HSC/MPPs had greater RNA content, were larger, more granular, and had lower CD49f protein expression than *TET2*^{WT} cells (Figure S6G).

Collectively, these data suggest CH^{MUT} HSC/MPPs have an altered transcriptional response to the inflammatory CH environment. To generate molecular signatures that capture differences

(D) GSEA against NF-κB, interferon, and hematopoietic signatures comparing *DNMT3A*^{MUT} versus *DNMT3A*^{WT} HSC/MPPs (left) and *TET2*^{MUT} versus *TET2*^{WT} HSC/MPPs (right) within CH samples. Signatures with FDR > 0.2 are colored gray. Positive NES values indicate enrichment in mutant cells. CB, cord blood.
(E) Heatmap showing $\log_2$ fold change in expression of genes related to inflammatory pathways within HSC/MPP. Left: comparison of CH^{WT} versus non-CH cells. Right: comparison of CH^{MUT} versus CH^{WT} cells within CH samples. Symbols represent FDR-corrected *p* values from differential expression testing.
(F) Strategy for deriving CH^{WT} and non-CH HSC/MPP signatures. Differential expression analysis was performed between HSC/MPPs from the 4 non-CH samples (n = 1,279 cells) and CH^{WT} cells from the 9 CH samples (n = 2,622 cells), accounting for sample, age, and batch effects.
(G) GSEA enrichment plots for the CH^{WT} HSC/MPP signature (top) and non-CH HSC/MPP signature (bottom), comparing CH^{WT} and CH^{MUT} cells within CH samples. Positive enrichment scores indicate enrichment in mutant cells.
See also Figure S6.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

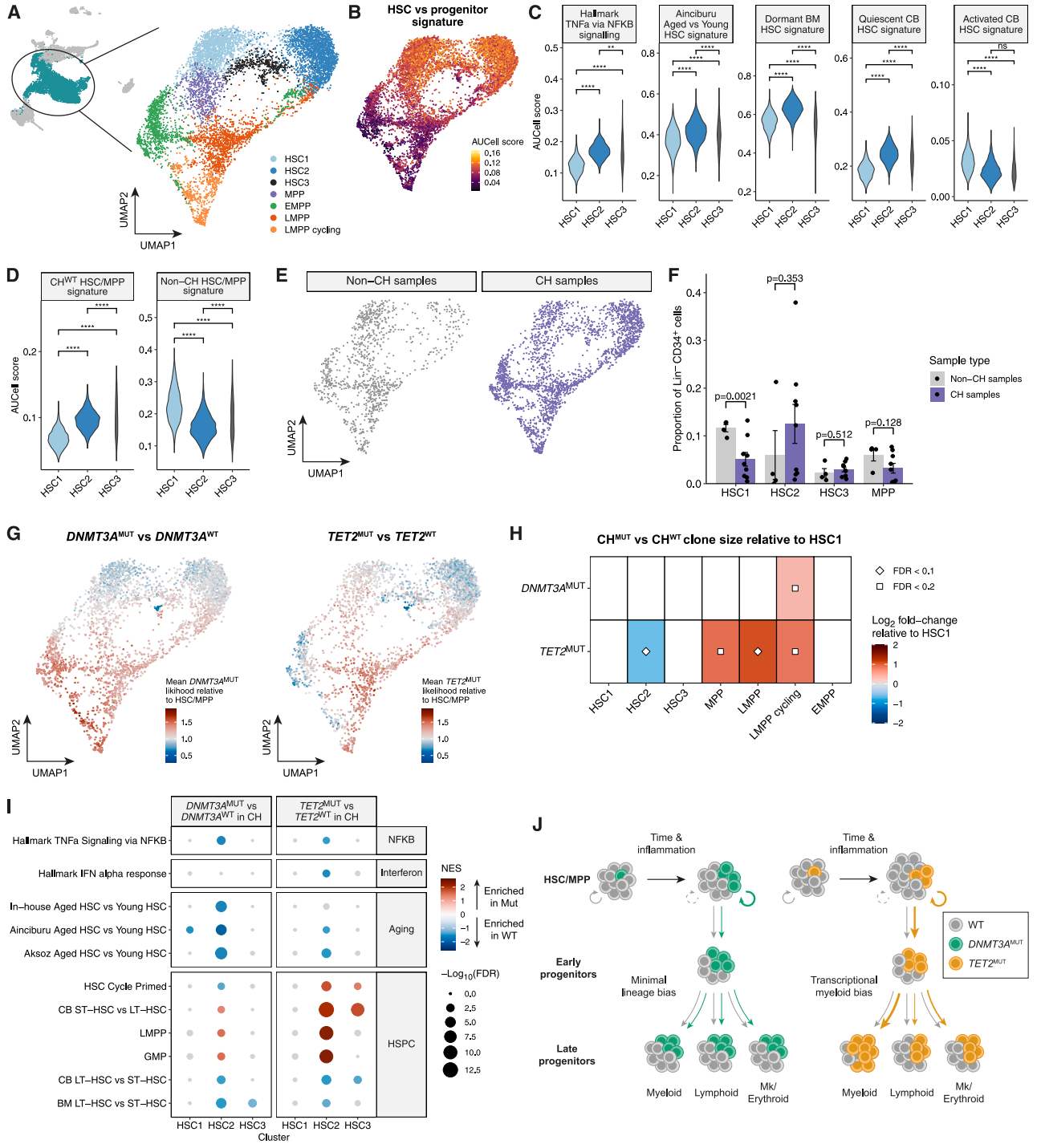


Figure 6. The effects of *DNMT3A* and *TET2* mutations are most prominent in a transcriptionally distinct subset of HSCs

(A) UMAP of 8,059 cells from the HSC/MPP, EMPP, LMPP, and LMPP cycling clusters after feature weight derivation with the SAM algorithm, colored by cluster annotation.

(B) UMAP superimposed with AUCell scores for a signature of genes differentially expressed between HSCs and progenitors.[66]

(C) AUCell scores for TNF-α via NF-κB, HSC aging, dormant BM HSC,[61] and quiescent versus activated CB HSC signatures,[89] comparing the HSC clusters. *p* values calculated by unpaired t test. The area of each violin is proportional to cell number.

(D) AUCell scores for the CH^{WT} and non-CH HSC/MPP signatures, comparing the HSC clusters. *p* values calculated by unpaired t test.

(E) UMAP embeddings showing cells from non-CH and CH samples.

(legend continued on next page)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

between the CH and non-CH environment, we compared gene expression profiles of CH^{WT} to non-CH HSC/MPPs (Figures 5F and S6H). Genes upregulated in CH^{WT} HSC/MPPs were enriched for TNF-α via NF-κB, HSC quiescence, and aging signatures (Figure S6I). We then asked whether *DNMT3A*^{MUT} and *TET2*^{MUT} HSC/MPPs were impacted by the CH environment differently than CH^{WT} HSC/MPPs. Indeed, both *DNMT3A*^{MUT} and *TET2*^{MUT} HSC/MPPs were negatively enriched for the CH^{WT} HSC/MPP signature (Figure 5G top), while *TET2*^{MUT} HSC/MPPs were positively enriched for the non-CH HSC/MPP signature (Figure 5G bottom), consistent with our findings above.

In summary, our data indicate that, compared with their CH^{WT} counterparts, CH^{MUT} HSC/MPPs are transcriptionally more similar to non-CH HSC/MPPs and have an attenuated transcriptional response to aging and inflammation.

The effects of *DNMT3A* and *TET2* mutations are most prominent in a transcriptionally distinct subset of HSCs

Transcriptional differences between CH and non-CH HSC/MPPs warranted a more granular exploration of HSC/MPP heterogeneity. Thus, we subclustered HSC/MPPs and the earliest progenitors from CH and non-CH samples (Figures 6A and S6J) using self-assembling manifolds (SAMs), an unsupervised approach to prioritize biologically relevant features among comparatively homogeneous cells, which has previously been applied to human HSC/MPPs.[89–91] This identified three distinct HSC clusters (HSC1–3), of which HSC1 and HSC2 contained most cells. All three HSC clusters showed transcriptional and immunophenotypic features of HSCs, although HSC3 was immunophenotypically more like MPPs (Figures 6B, S6K, and S6L).

Next, we asked whether the HSC clusters were transcriptionally distinct with respect to gene signatures of inflammation and aging (Figure 6C). Interestingly, HSC2 expressed higher levels of TNF-α/NF-κB signaling, aged HSC, and quiescence signatures,[61,89] and lower levels of the activated HSC signature. Concordant with their more quiescent transcriptional profile, HSC2 expressed fewer genes than HSC1 (Figure S6M). HSC3 exhibited a more heterogeneous expression profile, intermediate to HSC1 and HSC2. This was mirrored by the expression of exemplar genes associated with inflammatory signaling, quiescence, and cell cycle (Figure S6N). For example, HSC2 expressed lower levels of *CDK6*, which promotes exit from quiescence in LT-HSCs,[92] and higher levels of *GPRC5C*, which marks dormant human BM HSCs.[61] Interestingly, though HSC1 cells showed higher expression of genes that promote exit from quiescence, they also showed higher expression of genes and TF regulons implicated in HSC self-renewal (Figures S6N and S6O). Taken together, this suggests the HSC2 cluster has a transcriptional phenotype reflecting greater NF-κB pathway activity, increased quiescence, and decreased expression of TFs that support self-renewal.

Interestingly, HSC2 cells expressed higher levels of the CH^{WT} HSC/MPP signature and lower levels of the non-CH HSC/MPP signature (Figure 6D). To determine whether the heightened expression of inflammatory signatures in CH HSC/MPPs correlated with differences in HSC cluster composition, we compared the frequency of cells in the different HSC and MPP clusters. Interestingly, the HSC1 cluster was smaller in CH samples and the HSC2 cluster was larger, though the latter did not reach statistical significance (Figures 6E and 6F). This suggests differences in HSC cluster composition between CH and non-CH samples may provide a basis for the heightened transcriptional response to inflammation in CH.

The co-existence of distinct HSC states prompted us to ask whether CH mutations exerted context-specific transcriptional effects within the heterogeneous HSC compartment. Specifically, we sought to identify the cellular basis of attenuated expression of inflammatory and aging signatures in CH^{MUT} HSCs. One possible hypothesis is that CH^{MUT} cells are enriched in HSC1 compared with the HSC2 cluster. Examining the ratio of CH^{MUT} to CH^{WT} cells showed *TET2*^{MUT} cells were significantly depleted in HSC2 relative to HSC1, with clone size being smaller in 3 out of 4 *TET2*^{MUT} samples (Figures 6G, 6H, S6P, and S6Q). However, there was no significant difference for *DNMT3A*^{MUT} clones. Thus, depletion of mutant cells in HSC2 may partly account for the attenuated effect of inflammatory and aging programs in *TET2*^{MUT} HSCs.

A second, not mutually exclusive hypothesis, is that the transcriptional profiles of different HSC clusters could be differently modified by CH mutations. Specifically within HSC2, *DNMT3A*^{MUT} and *TET2*^{MUT} cells were negatively enriched for TNF-α signaling via NF-κB, aged HSC, and LT-HSC signatures, but positively enriched for ST-HSC, LMPP, and GMP signatures, relative to CH^{WT} HSCs (Figure 6I). Some of these differences were shared in the HSC1 and HSC3 clusters but were less marked. These data support the hypothesis that *DNMT3A* and *TET2* mutations, either directly or indirectly, attenuate expression of transcriptional programs related to inflammatory signaling and aging while

(F) Quantification of the size of each HSC/MPP cluster as a proportion of Lin^-CD34^+ cells, comparing CH and non-CH samples. Only cells sorted from the total Lin^-CD34^+ FACS gate were included. Data are represented as mean ± SEM. Each dot represents a sample. *p* values calculated by unpaired t test.

(G) UMAPs of cells from *DNMT3A*^{MUT} CH samples (left) and *TET2*^{MUT} CH samples (right) colored by the mean likelihood of cells being in the mutant clone relative the average in the HSC/MPP. The mean relative likelihood across all samples analyzed is shown ($n = 6$ *DNMT3A*^{MUT} samples; $n = 4$ *TET2*^{MUT} samples). A relative likelihood ≥ 1 indicates that the probability of a cell being mutant is greater than the average for the HSC/MPP.

(H) Heatmap of Log_2FC in abundance of mutant clones relative to the HSC1 cluster (from scCODA). Only significant results at FDR < 0.2 are shown; nonsignificant differences plotted as white.

(I) GSEA against NF-κB, interferon, aging, and hematopoietic signatures comparing *DNMT3A*^{MUT} versus *DNMT3A*^{WT} HSC/MPPs (left) and *TET2*^{MUT} versus *TET2*^{WT} HSC/MPPs (right) within CH samples. Signatures with FDR > 0.05 are colored gray. Positive NES values indicate enrichment in mutant cells.

(J) Model of *DNMT3A*^{MUT} and *TET2*^{MUT} clonal expansion. Inflammation impairs the function of CH^{WT} HSCs, but mutant HSCs are less affected, leading to clonal expansion over time (circular arrows). Downstream of the HSC, both *DNMT3A*^{MUT} and *TET2*^{MUT} clones expand moderately in early progenitors (linear arrows). In later differentiation, *DNMT3A*^{MUT} clone size is largely maintained, but *TET2*^{MUT} clones expand further and have a myeloid bias.

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

See also Figure S6.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

promoting lympho-myeloid differentiation programs, principally in the HSC2 cluster.

DISCUSSION

The first detailed single-cell examination of CH from older individuals with unperturbed hematopoiesis has uncovered new insights into the cell-intrinsic and cell-extrinsic consequences of CH mutations in humans. Our observations reflect the outcome of decades-long clonal competition within the human BM. The high-fidelity genotyping achieved with TARGET-seq+ ensures more than 90% of cells are genotyped, compared with ~20% with droplet-based approaches,[54] and is coupled to high-quality scRNA-seq data. Therefore, we can more confidently assign transcriptomes to either CH^{WT} or CH^{MUT} cells and discriminate clone-specific transcriptional programs.

Our data show that clones harboring heterozygous *DNMT3A* or *TET2* mutations follow a normal differentiation trajectory, although some lineage bias is observed. Quantitatively, most of the steady-state fitness advantage of $DNMT3A^{MUT}$ and $TET2^{MUT}$ clones occurs at the HSC/MPP level (Figure 6J). $DNMT3A^{MUT}$ clones show no consistent lineage bias, in contrast to data from myeloma remission samples, in which $DNMT3A^{R882}$-mutant CH cells were biased toward the megakaryocytic-erythroid lineage.[54] This discrepancy may either be due to true biological differences between $DNMT3A^{R882}$ and other *DNMT3A* mutations or to technical differences between TARGET-seq+ and droplet-based methods, the latter of which rely on RNA expression for genotyping. $TET2^{MUT}$ clones further expand 2- to 4-fold within the LMPP/GMP. Immature $TET2^{MUT}$ lympho-myeloid progenitors retain less HSC and non-myeloid transcriptional programs while showing premature upregulation of mature programs, possibly due to premature activity of myeloid transcription factors. These data are concordant with recent findings that $TET2^{MUT}$ HSPCs show exacerbated production of aberrant neutrophils in a cord blood knockout model.[56]

Importantly, our data support the notion of a more inflammatory environment in CH. Chronic inflammatory stimuli impair HSC self-renewal and mimic changes observed in aged mice.[93–96] Prior murine studies show that $Tet2^{-/-}$, $Dnmt3a^{-/-}$, and $Dnmt3a^{R878H/+}$ HSPCs exhibit clonal advantage in inflammatory environments.[45–50,97] Intestinal bacterial translocation,[48] chronic mycobacterial infection[45] or exposure to various cytokines, including IL-1,[49] IL-6,[47,48] TNF-α,[46,97] and IFN-γ,[45] promote clonal advantage of mutant HSPCs. Furthermore, Tet2 directly represses the pro-inflammatory cytokine IL-6.[32] However, no study has yet fully elucidated the differential impact of inflammation on co-existing WT and CH^{MUT} HSCs in native hematopoiesis. Furthermore, animal models may not adequately replicate human CH, where clones are exposed to changing environments over decades under heterogeneous selective pressures.

Our results suggest that CH mutations, directly or indirectly, attenuate the deleterious response to inflammation within HSCs. In agreement with our findings, data from zebrafish show that subclonal CH mutations lead to the expression of pro-inflammatory genes in mutant mature myeloid cells but anti-inflammatory genes in mutant HSPCs, providing them with a relative fitness advantage.[98] Furthermore, in mice, expansion of $Tet2^{+/-}$ subclones exposed to IL-1 is associated with attenuated downregulation of HSC self-renewal genes compared with WT HSCs,[49] whereas clonal advantage of $Dnmt3a^{R878H}$ HSPCs correlates with reduced activation of necroptosis in response to TNF-α challenge.[97] Together, these data suggest a model whereby the CH^{MUT} HSC population gains an advantage over the vastly more numerous non-mutant HSCs through differential impact of inflammation on their function (Figure 6J).

Further work is now needed to define the diverse drivers of inflammation in humans and the mechanisms by which human CH^{MUT} HSCs resist the detrimental effects of inflammation. The inflammatory drivers are likely to be heterogeneous and vary over time. Nevertheless, there may be common pathways of gene regulation downstream of multiple inflammatory signals. Examination of the BM niche in individuals with CH would thus be helpful. This, if combined with functional analyses and inferred rates of clonal expansion from the same individuals, may help to identify the most potent inflammatory drivers of CH^{MUT} clonal selection. These datasets could ultimately lead to therapies that diminish the inflammatory selective pressures or the ability of CH^{MUT} HSCs to resist the deleterious effects of inflammation and aging.

Limitations of the study

While we provide a hypothesis for clonal selection in human CH, functional experiments are required to define those pathways causative of clonal expansion, although prior functional murine studies support our findings.[32,45–47,49,97] Second, additional functions of DNMT3A and TET2 beyond transcriptional regulation[30–37] may also play important roles in clonal advantage. Third, our findings do not fully reveal the mechanistic link between altered DNMT3A/TET2 function and downstream pathways. Future studies integrating single-cell genotyping with chromatin accessibility[99] and DNA binding in primary CH samples may elucidate the epigenetic basis for the observed transcriptional changes.

Since osteoarthritis is associated with low-grade chronic inflammation,[100] the samples analyzed may have been more impacted by inflammation than in individuals without osteoarthritis. However, since both CH and non-CH individuals had osteoarthritis, all our analyses are controlled for this effect.

Importantly, we appreciate that our conclusions depend on the accuracy of single-cell genotyping. All single-cell genotyping approaches are affected by ADO, which varies considerably depending on the method. ADO can result in significant mis-genotyping, impacting downstream analyses, especially when clones are small. Importantly, TARGET-seq+ yields lower ADO rates compared with droplet-based technologies. In our study, cells in which ADO of the mutant allele occurred could only be identified at four loci where a heterozygous germline SNP was present in the amplicon (Figures S3E–S3H). We estimate the ADO rate to be 5%–15%, depending on the locus and quantification method used (STAR Methods). For the other 8 mutations, the inferred ADO rate was approximately 5%–13.5% (Figures S3I and S3J). This would have resulted in 5%–13.5% of mutant cells being called WT. We have tested the impact of this ADO rate on downstream analyses by performing simulations in which we deliberately mis-classified 10% of mutant HSC/MPP cells as WT. The GSEA results obtained were highly concordant with the original analysis, albeit with reduced enrichment scores and significance

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

(Table S6). Overall, this suggests that modest ADO rates lead to a subtle "dilution" of the apparent differences between CH^{WT} and CH^{MUT} cells, without altering our conclusions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Patient samples
- METHOD DETAILS
  - Sample collection and processing
  - Targeted DNA sequencing
  - Flow cytometry and FACS sorting
  - Whole-genome sequencing (WGS) analysis
  - Droplet digital PCR (ddPCR)
  - TARGET-seq+ library preparation
  - TARGET-seq+ validation experiments
  - Targeted single-cell genotyping analysis
  - Analysis of FACS index data
  - Single-cell transcriptome data pre-processing
  - Single-cell transcriptome analysis
  - FACS sorting and snRNA-seq for 'in-house' aging dataset
  - Single-nucleus RNA-seq processing - In-house aging dataset
  - Single-cell RNA-seq processing - Ainciburu et al. dataset
  - Single-cell RNA-seq processing - Aksöz et al. dataset
  - Aged vs Young HSC Differential Expression
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2024.05.010.

ACKNOWLEDGMENTS

P.V. acknowledges funding from the Medical Research Council Molecular Haematology Unit Programme Grant (MC_UU_00029/8), Blood Cancer UK Programme Continuity Grant 13008, NIHR Senior Fellowship, and the Oxford Biomedical Research Centre Haematology Theme. N.A.J. was supported by a Medical Research Council and Leukaemia UK Clinical Research Training Fellowship (MR/R002258/1). S.T. is supported by a Scatcherd European Scholarship in partnership with the Medical Research Council, the Radcliffe Department of Medicine, and the Clarendon Fund. A.G.X.Z. is supported by a University of Toronto MD/PhD studentship award. M.M., R.M., B.U., M.A.S., and A.K. were funded by the Haematology Theme of the Oxford NIHR Biomedical Research Centre. Work in the laboratory of J.E.D. is supported by funds from the Princess Margaret Cancer Foundation, the Ontario Institute for Cancer Research through funding provided by the Government of Ontario, the Canadian Institutes for Health Research (RN380110-409786), the International Development Research Centre, Ottawa, Canada, the Canadian Cancer Society (703212), a Terry Fox New Frontiers Program project grant, the University of Toronto's Medicine by Design initiative with funding from the Canada First Research Excellence Fund, the Ontario Ministry of Health, and a Canada Research Chair. S.G.D. is funded by a Versus Arthritis Career Development Fellowship (22425). We thank Prof. Doug Higgs, Prof. Thomas Höfer, Dr. Verena Körber, Dr. David Cruz Hernandez, and Mr. Angus Groom for insightful comments and discussions. The authors acknowledge the MRC WIMM Flow Cytometry and Single Cell Facilities. Some figures were created using BioRender.

AUTHOR CONTRIBUTIONS

Conceptualization, N.A.J., S.T., and P.V.; methodology, N.A.J., S.T., and A.G.X.Z.; investigation, N.A.J., S.T., B.S., M.M., S.R., M.S.N., S.S., R.M., A.K., A.G.A.-N., and E.F.-F.; formal analysis, N.A.J., A.G.X.Z., S.S., G.-A.G., and A.M.; visualization, N.A.J., A.G.X.Z., and S.S.; resources, R.M., B.U., M.A.S., S.N., B.J.L.K., A.H.T., R.A.-L., R.G., A.G.A.-N., and E.F.-F.; project administration, B.W., K.W., and D.B.; supervision, S.G.D., A.J.C., C.N., J.E.D., S.Z.X., and P.V.; funding acquisition, N.A.J., J.E.D., and P.V.; writing – original draft, N.A.J., S.T., A.G.X.Z., J.E.D., S.Z.X., and P.V.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

J.E.D. receives royalties from Trillium Therapeutics Inc./Pfizer and a commercial research grant from Celgene/BMS.

Received: June 25, 2023
Revised: January 29, 2024
Accepted: May 30, 2024
Published: June 24, 2024

REFERENCES

1. Welch, J.S., Ley, T.J., Link, D.C., Miller, C.A., Larson, D.E., Koboldt, D.C., Wartman, L.D., Lamprecht, T.L., Liu, F., Xia, J., et al. (2012). The origin and evolution of mutations in acute myeloid leukemia. Cell *150*, 264–278. https://doi.org/10.1016/j.cell.2012.06.023.
2. Weissman, I.L. (2015). Stem cells are units of natural selection for tissue formation, for germline development, and in cancer development. Proc. Natl. Acad. Sci. USA *112*, 8922–8928. https://doi.org/10.1073/pnas.1505464112.
3. Ziegler, A., Jonason, A.S., Leffell, D.J., Simon, J.A., Sharma, H.W., Kimmelman, J., Remington, L., Jacks, T., and Brash, D.E. (1994). Sunburn and p53 in the onset of skin cancer. Nature *372*, 773–776. https://doi.org/10.1038/372773a0.
4. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science *348*, 880–886. https://doi.org/10.1126/science.aaa6806.
5. Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. Nature *538*, 260–264. https://doi.org/10.1038/nature19768.
6. Martincorena, I., Fowler, J.C., Wabik, A., Lawson, A.R.J., Abascal, F., Hall, M.W.J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M.R., et al. (2018). Somatic mutant clones colonize the human esophagus with age. Science *362*, 911–917. https://doi.org/10.1126/science.aau3879.
7. Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., Shiozawa, Y., Sato, Y., Aoki, K., Kim, S.K., et al. (2019). Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature *565*, 312–317. https://doi.org/10.1038/s41586-018-0811-x.
8. Yizhak, K., Aguet, F., Kim, J., Hess, J.M., Kübler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N.J., Rosebrock, D., et al. (2019). RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. Science *364*, eaaw0726. https://doi.org/10.1126/science.aaw0726.
9. Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R.J., Sanders, M.A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., et al. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. Nature *574*, 532–537. https://doi.org/10.1038/s41586-019-1672-7.
10. Yoshida, K., Gowers, K.H.C., Lee-Six, H., Chandrasekharan, D.P., Coorens, T., Maughan, E.F., Beal, K., Menzies, A., Millar, F.R., Anderson, E., et al. (2020). Tobacco smoking and somatic mutations in

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

human bronchial epithelium. Nature *578*, 266–272. https://doi.org/10.1038/s41586-020-1961-1.

11. Nanki, K., Fujii, M., Shimokawa, M., Matano, M., Nishikori, S., Date, S., Takano, A., Toshimitsu, K., Ohta, Y., Takahashi, S., et al. (2020). Somatic inflammatory gene mutations in human ulcerative colitis epithelium. Nature *577*, 254–259. https://doi.org/10.1038/s41586-019-1844-5.
12. Kakiuchi, N., Yoshida, K., Uchino, M., Kihara, T., Akaki, K., Inoue, Y., Kawada, K., Nagayama, S., Yokoyama, A., Yamamoto, S., et al. (2020). Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. Nature *577*, 260–265. https://doi.org/10.1038/s41586-019-1856-1.
13. Busque, L., Patel, J.P., Figueroa, M.E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A., et al. (2012). Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat. Genet. *44*, 1179–1181. https://doi.org/10.1038/ng.2413.
14. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. *371*, 2488–2498. https://doi.org/10.1056/NEJMoa1408617.
15. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. *371*, 2477–2487. https://doi.org/10.1056/NEJMoa1409405.
16. Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. *20*, 1472–1478. https://doi.org/10.1038/nm.3733.
17. Desai, P., Mencia-Trinchant, N., Savenkov, O., Simon, M.S., Cheang, G., Lee, S., Samuel, M., Ritchie, E.K., Guzman, M.L., Ballman, K.V., et al. (2018). Somatic mutations precede acute myeloid leukemia years before diagnosis. Nat. Med. *24*, 1015–1023. https://doi.org/10.1038/s41591-018-0081-z.
18. Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., et al. (2017). Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. N. Engl. J. Med. *377*, 111–121. https://doi.org/10.1056/NEJMoa1701719.
19. Abelson, S., Collord, G., Ng, S.W.K., Weissbrod, O., Mendelson Cohen, N., Niemeyer, E., Barda, N., Zuzarte, P.C., Heisler, L., Sundaravadanam, Y., et al. (2018). Prediction of acute myeloid leukaemia risk in healthy individuals. Nature *559*, 400–404. https://doi.org/10.1038/s41586-018-0317-6.
20. Dorsheimer, L., Assmus, B., Rasper, T., Ortmann, C.A., Ecke, A., Abou-El-Ardat, K., Schmid, T., Brüne, B., Wagner, S., Serve, H., et al. (2019). Association of mutations contributing to clonal hematopoiesis with prognosis in chronic ischemic heart failure. JAMA Cardiol. *4*, 25–33. https://doi.org/10.1001/jamacardio.2018.3965.
21. Kessler, M.D., Damask, A., O'Keeffe, S., Banerjee, N., Li, D., Watanabe, K., Marketta, A., Van Meter, M., Semrau, S., Horowitz, J., et al. (2022). Common and rare variant associations with clonal haematopoiesis phenotypes. Nature *612*, 301–309. https://doi.org/10.1038/s41586-022-05448-9.
22. Wong, W.J., Emdin, C., Bick, A.G., Zekavat, S.M., Niroula, A., Pirruccello, J.P., Dichtel, L., Griffin, G., Uddin, M.M., Gibson, C.J., et al. (2023). Clonal haematopoiesis and risk of chronic liver disease. Nature *616*, 747–754. https://doi.org/10.1038/s41586-023-05857-4.
23. Skead, K., Ang Houle, A., Abelson, S., Agbessi, M., Bruat, V., Lin, B., Soave, D., Shlush, L., Wright, S., Dick, J., et al. (2021). Interacting evolutionary pressures drive mutation dynamics and health outcomes in aging blood. Nat. Commun. *12*, 4921. https://doi.org/10.1038/s41467-021-25172-8.
24. Bick, A.G., Weinstock, J.S., Nandakumar, S.K., Fulco, C.P., Bao, E.L., Zekavat, S.M., Szeto, M.D., Liao, X., Leventhal, M.J., Nasser, J., et al. (2020). Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature *586*, 763–768. https://doi.org/10.1038/s41586-020-2819-2.
25. Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6.
26. Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133. https://doi.org/10.1038/nature09303.
27. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303. https://doi.org/10.1126/science.1210597.
28. Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930. https://doi.org/10.1126/science.1169786.
29. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell *151*, 1417–1430. https://doi.org/10.1016/j.cell.2012.11.022.
30. Fuks, F., Burgers, W.A., Godin, N., Kasai, M., and Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J. *20*, 2536–2544. https://doi.org/10.1093/emboj/20.10.2536.
31. Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. Nature *493*, 561–564. https://doi.org/10.1038/nature11742.
32. Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., Zhao, D., Liu, Y., Wang, C., Zhang, X., et al. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature *525*, 389–393. https://doi.org/10.1038/nature15252.
33. Shen, Q., Zhang, Q., Shi, Y., Shi, Q., Jiang, Y., Gu, Y., Li, Z., Li, X., Zhao, K., Wang, C., et al. (2018). Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. Nature *554*, 123–127. https://doi.org/10.1038/nature25434.
34. Guallar, D., Bi, X., Pardavila, J.A., Huang, X., Saenz, C., Shi, X., Zhou, H., Faiola, F., Ding, J., Haruehanroengra, P., et al. (2018). RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. Nat. Genet. *50*, 443–451. https://doi.org/10.1038/s41588-018-0060-9.
35. Lan, J., Rajan, N., Bizet, M., Penning, A., Singh, N.K., Guallar, D., Calonne, E., Li Greci, A., Bonvin, E., Deplus, R., et al. (2020). Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. Nat. Commun. *11*, 4956. https://doi.org/10.1038/s41467-020-18729-6.
36. He, C., Bozler, J., Janssen, K.A., Wilusz, J.E., Garcia, B.A., Schorn, A.J., and Bonasio, R. (2021). TET2 chemically modifies tRNAs and regulates tRNA fragment levels. Nat. Struct. Mol. Biol. *28*, 62–70. https://doi.org/10.1038/s41594-020-00526-w.
37. Ramabadran, R., Wang, J.H., Reyes, J.M., Guzman, A.G., Gupta, S., Rosas, C., Brunetti, L., Gundry, M.C., Tovy, A., Long, H., et al. (2023). DNMT3A-coordinated splicing governs the stem state switch towards differentiation in embryonic and haematopoietic stem cells. Nat. Cell Biol. *25*, 528–539. https://doi.org/10.1038/s41556-023-01109-9.
38. Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat. Genet. *44*, 23–31. https://doi.org/10.1038/ng.1009.
39. Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

homeostasis and differentiation of hematopoietic stem cells in mice. Proc. Natl. Acad. Sci. USA *108*, 14566–14571. https://doi.org/10.1073/pnas.1112317108.

40. Quivoron, C., Couronné, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.-H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell *20*, 25–38. https://doi.org/10.1016/j.ccr.2011.06.003.

41. Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). TET2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell *20*, 11–24. https://doi.org/10.1016/j.ccr.2011.06.001.

42. Li, Z., Cai, X., Cai, C.-L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.-C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509–4518. https://doi.org/10.1182/blood-2010-12-325241.

43. Ostrander, E.L., Kramer, A.C., Mallaney, C., Celik, H., Koh, W.K., Fairchild, J., Haussler, E., Zhang, C.R.C., and Challen, G.A. (2020). Divergent effects of Dnmt3a and Tet2 mutations on hematopoietic progenitor cell fitness. Stem Cell Rep. *3*, 539–547. https://doi.org/10.1016/j.stemcr.2020.02.011.

44. Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat. Genet. *52*, 378–387. https://doi.org/10.1038/s41588-020-0595-4.

45. Hormaechea-Agulla, D., Matatall, K.A., Le, D.T., Kain, B., Long, X., Kus, P., Jaksik, R., Challen, G.A., Kimmel, M., and King, K.Y. (2021). Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling. Cell Stem Cell *28*, 1428–1442.e6. https://doi.org/10.1016/j.stem.2021.03.002.

46. SanMiguel, J.M., Eudy, E., Loberg, M.A., Young, K.A., Mistry, J.J., Mujica, K.D., Schwartz, L.S., Stearns, T.M., Challen, G.A., and Trowbridge, J.J. (2022). Distinct tumor necrosis factor alpha Receptors Dictate Stem Cell Fitness versus Lineage Output in Dnmt3a-Mutant Clonal Hematopoiesis. Cancer Discov. *12*, 2763–2773. https://doi.org/10.1158/2159-8290.CD-22-0086.

47. Cai, Z., Kotzin, J.J., Ramdas, B., Chen, S., Nelanuthala, S., Palam, L.R., Pandey, R., Mali, R.S., Liu, Y., Kelley, M.R., et al. (2018). Inhibition of inflammatory signaling in Tet2 mutant preleukemic cells mitigates stress-induced abnormalities and clonal hematopoiesis. Cell Stem Cell *23*, 833–849.e5. https://doi.org/10.1016/j.stem.2018.10.013.

48. Meisel, M., Hinterleitner, R., Pacis, A., Chen, L., Earley, Z.M., Mayassi, T., Pierre, J.F., Ernest, J.D., Galipeau, H.J., Thuille, N., et al. (2018). Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. Nature *557*, 580–584. https://doi.org/10.1038/s41586-018-0125-z.

49. Caiado, F., Kovtonyuk, L.V., Gonullu, N.G., Fullin, J., Boettcher, S., and Manz, M.G. (2023). Aging drives Tet2+/- clonal hematopoiesis via IL-1 signaling. Blood *141*, 886–903. https://doi.org/10.1182/blood.2022016835.

50. Florez, M.A., Tran, B.T., Wathan, T.K., DeGregori, J., Pietras, E.M., and King, K.Y. (2022). Clonal hematopoiesis: mutation-specific adaptation to environmental change. Cell Stem Cell *29*, 882–904. https://doi.org/10.1016/j.stem.2022.05.006.

51. Fabre, M.A., de Almeida, J.G., Fiorillo, E., Mitchell, E., Damaskou, A., Rak, J., Orrù, V., Marongiu, M., Chapman, M.S., Vijayabaskar, M.S., et al. (2022). The longitudinal dynamics and natural history of clonal haematopoiesis. Nature *606*, 335–342. https://doi.org/10.1038/s41586-022-04785-z.

52. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. Nature *606*, 343–350. https://doi.org/10.1038/s41586-022-04786-y.

53. Watson, C.J., Papula, A.L., Poon, G.Y.P., Wong, W.H., Young, A.L., Druley, T.E., Fisher, D.S., and Blundell, J.R. (2020). The evolutionary dynamics and fitness landscape of clonal hematopoiesis. Science *367*, 1449–1454. https://doi.org/10.1126/science.aay9333.

54. Nam, A.S., Dusaj, N., Izzo, F., Murali, R., Myers, R.M., Mouhieddine, T.H., Sotelo, J., Benbarche, S., Waarts, M., Gaiti, F., et al. (2022). Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. Nat. Genet. *54*, 1514–1526. https://doi.org/10.1038/s41588-022-01179-9.

55. Nakauchi, Y., Azizi, A., Thomas, D., Corces, M.R., Reinisch, A., Sharma, R., Cruz Hernandez, D., Köhnke, T., Karigane, D., Fan, A., et al. (2022). The cell type–specific 5hmC landscape and dynamics of healthy human hematopoiesis and TET2-mutant preleukemia. Blood Cancer Discovery, bloodcandisc.BCD-A.2021. Blood Cancer Discov. *3*, 346–367. https://doi.org/10.1158/2643-3230.BCD-21-0143.

56. Huerga Encabo, H., Aramburu, I.V., Garcia-Albornoz, M., Piganeau, M., Wood, H., Song, A., Ferrelli, A., Sharma, A., Minutti, C.M., Domart, M.-C., et al. (2023). Loss of TET2 in human hematopoietic stem cells alters the development and function of neutrophils. Cell Stem Cell *30*, 781–799.e9. https://doi.org/10.1016/j.stem.2023.05.004.

57. Hecker, J.S., Hartmann, L., Rivière, J., Buck, M.C., van der Garde, M., Rothenberg-Thurley, M., Fischer, L., Winter, S., Ksienzyk, B., Ziemann, F., et al. (2021). CHIP and hips: clonal hematopoiesis is common in patients undergoing hip arthroplasty and is associated with autoimmune disease. Blood *138*, 1727–1732. https://doi.org/10.1182/blood.2020010163.

58. van Zeventer, I.A., de Graaf, A.O., Salzbrunn, J.B., Nolte, I.M., Kamphuis, P., Dinmohamed, A., van der Reijden, B.A., Schuringa, J.J., Jansen, J.H., and Huls, G. (2023). Evolutionary landscape of clonal hematopoiesis in 3,359 individuals from the general population. Cancer Cell *41*, 1017–1031.e4. https://doi.org/10.1016/j.ccell.2023.04.006.

59. Rodriguez-Meira, A., Buck, G., Clark, S.-A., Povinelli, B.J., Alcolea, V., Louka, E., McGowan, S., Hamblin, A., Sousos, N., Barkas, N., et al. (2019). Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. Mol. Cell *73*, 1292–1305.e8. https://doi.org/10.1016/j.molcel.2019.01.009.

60. Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A.J.M., Faridani, O.R., and Sandberg, R. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat. Biotechnol. *38*, 708–714. https://doi.org/10.1038/s41587-020-0497-0.

61. Zhang, Y.W., Mess, J., Aizarani, N., Mishra, P., Johnson, C., Romero-Mulero, M.C., Rettkowski, J., Schönberger, K., Obier, N., Jäcklein, K., et al. (2022). Hyaluronic acid–GPRC5C signalling promotes dormancy in haematopoietic stem cells. Nat. Cell Biol. *24*, 1038–1048. https://doi.org/10.1038/s41556-022-00931-x.

62. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281. https://doi.org/10.1038/ncb3493.

63. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309. https://doi.org/10.1016/j.cell.2011.01.004.

64. Drissen, R., Thongjuea, S., Theilgaard-Mönch, K., and Nerlov, C. (2019). Identification of two distinct pathways of human myelopoiesis. Sci. Immunol. *4*, eaau7148. https://doi.org/10.1126/sciimmunol.aau7148.

65. Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., Meira, A.R., et al. (2021). Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. Cell Rep. *36*, 109698. https://doi.org/10.1016/j.celrep.2021.109698.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

66. Xie, S.Z., Kaufmann, K.B., Wang, W., Chan-Seng-Yue, M., Gan, O.I., Laurenti, E., Garcia-Prat, L., Takayanagi, S.-I., Ng, S.W.K., Xu, C., et al. (2021). Sphingosine-1-phosphate Receptor 3 potentiates inflammatory programs in normal and leukemia stem cells to promote differentiation. Blood Cancer Discov. *2*, 32–53. https://doi.org/10.1158/2643-3230.BCD-20-0155.

67. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., et al. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nat. Immunol. *19*, 85–97. https://doi.org/10.1038/s41590-017-0001-2.

68. Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., et al. (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science *345*, 1251033. https://doi.org/10.1126/science.1251033.

69. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.

70. Drissen, R., Buza-Vidas, N., Woll, P., Thongjuea, S., Gambardella, A., Giustacchini, A., Mancini, E., Zriwil, A., Lutteropp, M., Grover, A., et al. (2016). Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. Nat. Immunol. *17*, 666–676. https://doi.org/10.1038/ni.3412.

71. Pang, W.W., Price, E.A., Sahoo, D., Beerman, I., Maloney, W.J., Rossi, D.J., Schrier, S.L., and Weissman, I.L. (2011). Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proc. Natl. Acad. Sci. USA *108*, 20012–20017. https://doi.org/10.1073/pnas.1116110108.

72. Sternberg, A., Killick, S., Littlewood, T., Hatton, C., Peniket, A., Seidl, T., Soneji, S., Leach, J., Bowen, D., Chapman, C., et al. (2005). Evidence for reduced B-cell progenitors in early (low-risk) myelodysplastic syndrome. Blood *106*, 2982–2991. https://doi.org/10.1182/blood-2005-04-1543.

73. Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R., et al. (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell *19*, 138–152. https://doi.org/10.1016/j.ccr.2010.12.012.

74. Quek, L., Otto, G.W., Garnett, C., Lhermitte, L., Karamitros, D., Stoilova, B., Lau, I.J., Doondeea, J., Usukhbayar, B., Kennedy, A., et al. (2016). Genetically distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. J. Exp. Med. *213*, 1513–1535. https://doi.org/10.1084/jem.20151775.

75. Woll, P.S., Kjällquist, U., Chowdhury, O., Doolittle, H., Wedge, D.C., Thongjuea, S., Erlandsson, R., Ngara, M., Anderson, K., Deng, Q., et al. (2014). Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. Cancer Cell *25*, 794–808. https://doi.org/10.1016/j.ccr.2014.03.036.

76. Büttner, M., Ostner, J., Müller, C.L., Theis, F.J., and Schubert, B. (2021). scCODA is a Bayesian model for compositional single-cell data analysis. Nat. Commun. *12*, 6876. https://doi.org/10.1038/s41467-021-27150-6.

77. Weinstock, J.S., Gopakumar, J., Burugula, B.B., Uddin, M.M., Jahn, N., Belk, J.A., Bouzid, H., Daniel, B., Miao, Z., Ly, N., et al. (2023). Aberrant activation of TCL1A promotes stem cell expansion in clonal haematopoiesis. Nature *616*, 755–763. https://doi.org/10.1038/s41586-023-05806-1.

78. Arends, C.M., Galan-Sousa, J., Hoyer, K., Chan, W., Jäger, M., Yoshida, K., Seemann, R., Noerenberg, D., Waldhueter, N., Fleischer-Notter, H., et al. (2018). Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. Leukemia *32*, 1908–1919. https://doi.org/10.1038/s41375-018-0047-7.

79. Buscarlet, M., Provost, S., Zada, Y.F., Bourgoin, V., Mollica, L., Dubé, M.-P., and Busque, L. (2018). Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. Blood *132*, 277–280. https://doi.org/10.1182/blood-2018-01-829937.

80. Thol, F., Klesse, S., Köhler, L., Gabdoulline, R., Kloos, A., Liebich, A., Wichmann, M., Chaturvedi, A., Fabisch, J., Gaidzik, V.I., et al. (2017). Acute myeloid leukemia derived from lympho-myeloid clonal hematopoiesis. Leukemia *31*, 1286–1295. https://doi.org/10.1038/leu.2016.345.

81. Tulstrup, M., Soerensen, M., Hansen, J.W., Gillberg, L., Needhamsen, M., Kaastrup, K., Helin, K., Christensen, K., Weischenfeldt, J., and Grønbæk, K. (2021). TET2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis. Nat. Commun. *12*, 6061. https://doi.org/10.1038/s41467-021-26093-2.

82. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

83. Lee-Six, H., Øbro, N.F., Shepherd, M.S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R.J., Huntly, B.J.P., Martincorena, I., Anderson, E., et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. Nature *561*, 473–478. https://doi.org/10.1038/s41586-018-0497-0.

84. Ainciburu, M., Ezponda, T., Berastegui, N., Alfonso-Pierola, A., Vilas-Zornoza, A., San Martin-Uriz, P., Alignani, D., Lamo-Espinosa, J., San-Julian, M., Jiménez-Solas, T., et al. (2023). Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. eLife *12*, e79363. https://doi.org/10.7554/eLife.79363.

85. Aksöz, M., Gafencu, G.-A., Stoilova, B., Buono, M., Meng, Y., Jakobsen, N.A., Metzner, M., Clark, S.-A., Beveridge, R., Thongjuea, S., et al. (2022). Identification and age-dependent increase of platelet biased human hematopoietic stem cells. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.14.475546.

86. Chambers, S.M., Shaw, C.A., Gatza, C., Fisk, C.J., Donehower, L.A., and Goodell, M.A. (2007). Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. PLoS Biol. *5*, e201. https://doi.org/10.1371/journal.pbio.0050201.

87. Hammond, C.A., Wu, S.W., Wang, F., MacAldaz, M.E., and Eaves, C.J. (2023). Aging alters the cell cycle control and mitogenic signaling responses of human hematopoietic stem cells. Blood *141*, 1990–2002. https://doi.org/10.1182/blood.2022017174.

88. Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M., and de Haan, G. (2011). Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J. Exp. Med. *208*, 2691–2703. https://doi.org/10.1084/jem.20111490.

89. García-Prat, L., Kaufmann, K.B., Schneiter, F., Voisin, V., Murison, A., Chen, J., Chan-Seng-Yue, M., Gan, O.I., McLeod, J.L., Smith, S.A., et al. (2021). TFEB-mediated endolysosomal activity controls human hematopoietic stem cell fate. Cell Stem Cell *28*, 1838–1850.e10. https://doi.org/10.1016/j.stem.2021.07.003.

90. Tarashansky, A.J., Xue, Y., Li, P., Quake, S.R., and Wang, B. (2019). Self-assembling manifolds in single-cell RNA sequencing data. eLife *8*, e48994. https://doi.org/10.7554/eLife.48994.

91. Xie, S.Z., Garcia-Prat, L., Voisin, V., Ferrari, R., Gan, O.I., Wagenblast, E., Kaufmann, K.B., Zeng, A.G.X., Takayanagi, S.-I., Patel, I., et al. (2019). Sphingolipid modulation activates proteostasis programs to govern human hematopoietic stem cell self-renewal. Cell Stem Cell *25*, 639–653.e7. https://doi.org/10.1016/j.stem.2019.09.008.

92. Laurenti, E., Frelin, C., Xie, S., Ferrari, R., Dunant, C.F., Zandi, S., Neumann, A., Plumb, I., Doulatov, S., Chen, J., et al. (2015). CDK6 levels regulate quiescence exit in human hematopoietic stem cells. Cell Stem Cell *16*, 302–313. https://doi.org/10.1016/j.stem.2015.01.017.

93. Pietras, E.M., Mirantes-Barbeito, C., Fong, S., Loeffler, D., Kovtonyuk, L.V., Zhang, S., Lakshminarasimhan, R., Chin, C.P., Techner, J.-M., Will, B., et al. (2016). Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

expense of self-renewal. Nat. Cell Biol. *18*, 607–618. https://doi.org/10.1038/ncb3346.

94. Matatall, K.A., Jeong, M., Chen, S., Sun, D., Chen, F., Mo, Q., Kimmel, M., and King, K.Y. (2016). Chronic infection depletes hematopoietic stem cells through stress-induced terminal differentiation. Cell Rep. *17*, 2584–2595. https://doi.org/10.1016/j.celrep.2016.11.031.
95. Bogeska, R., Mikecin, A.-M., Kaschutnig, P., Fawaz, M., Büchler-Schäff, M., Le, D., Ganuza, M., Vollmer, A., Paffenholz, S.V., Asada, N., et al. (2022). Inflammatory exposure drives long-lived impairment of hematopoietic stem cell self-renewal activity and accelerated aging. Cell Stem Cell *29*, 1273–1284.e8. https://doi.org/10.1016/j.stem.2022.06.012.
96. Mitchell, C.A., Verovskaya, E.V., Calero-Nieto, F.J., Olson, O.C., Swann, J.W., Wang, X., Hérault, A., Dellorusso, P.V., Zhang, S.Y., Svendsen, A.F., et al. (2023). Stromal niche inflammation mediated by IL-1 signalling is a targetable driver of haematopoietic ageing. Nat. Cell Biol. *25*, 30–41. https://doi.org/10.1038/s41556-022-01053-0.
97. Liao, M., Chen, R., Yang, Y., He, H., Xu, L., Jiang, Y., Guo, Z., He, W., Jiang, H., and Wang, J. (2022). Aging-elevated inflammation promotes DNMT3A R878H-driven clonal hematopoiesis. Acta Pharm. Sin. B *12*, 678–691. https://doi.org/10.1016/j.apsb.2021.09.015.
98. Avagyan, S., Henninger, J.E., Mannherz, W.P., Mistry, M., Yoon, J., Yang, S., Weber, M.C., Moore, J.L., and Zon, L.I. (2021). Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. Science *374*, 768–772. https://doi.org/10.1126/science.aba9304.
99. Turkalj, S., Jakobsen, N.A., Groom, A., Metzner, M., Riva, S.G., Gür, E.R., Usukhbayar, B., Salazar, M.A., Hentges, L.D., Mickute, G., et al. (2023). GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells. Cell Stem Cell *30*, 722–740.e11. https://doi.org/10.1016/j.stem.2023.04.012.
100. Robinson, W.H., Lepus, C.M., Wang, Q., Raghu, H., Mao, R., Lindstrom, T.M., and Sokolove, J. (2016). Low-grade inflammation as a key mediator of the pathogenesis of osteoarthritis. Nat. Rev. Rheumatol. *12*, 580–592. https://doi.org/10.1038/nrrheum.2016.136.
101. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat. Protoc. *15*, 2247–2276. https://doi.org/10.1038/s41596-020-0336-2.
102. Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.
103. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
104. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson, J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. *44*, e108. https://doi.org/10.1093/nar/gkw227.
105. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. *10*, 1556–1566. https://doi.org/10.1038/nprot.2015.105.
106. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115. https://doi.org/10.1093/nar/gks596.
107. Aguilar-Navarro, A.G., Meza-León, B., Gratzinger, D., Juárez-Aguilar, F.G., Chang, Q., Ornatsky, O., Tsui, H., Esquivel-Gómez, R., Hernández-Ramírez, A., Xie, S.Z., et al. (2020). Human aging alters the spatial organization between CD34+ hematopoietic cells and adipocytes in bone marrow. Stem Cell Rep. *15*, 317–325. https://doi.org/10.1016/j.stemcr.2020.06.011.
108. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.
109. Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S.H., Kiemeney, L.A.L.M., Veltman, J.A., Gilissen, C., and Hoischen, A. (2017). Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. Am. J. Hum. Genet. *101*, 50–64. https://doi.org/10.1016/j.ajhg.2017.05.013.
110. Körber, V., Ansari-Pour, N., Jakobsen, N.A., Moore, R., Claudino, N., Metzner, M., Hörsch, F., Usukhbayar, B., Salazar, M.A., Newman, S., et al. (2023). Detecting and quantifying clonal selection in somatic stem cells. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.15.472780.
111. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134. https://doi.org/10.1186/1471-2105-13-134.
112. Quek, L., David, M.D., Kennedy, A., Metzner, M., Amatangelo, M., Shih, A., Stoilova, B., Quivoron, C., Heiblig, M., Willekens, C., et al. (2018). Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib. Nat. Med. *24*, 1167–1177. https://doi.org/10.1038/s41591-018-0115-6.
113. Jahn, K., Kuipers, J., and Beerenwinkel, N. (2016). Tree inference for single-cell data. Genome Biol. *17*, 86. https://doi.org/10.1186/s13059-016-0936-x.
114. Lun, A.T.L., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. *17*, 75. https://doi.org/10.1186/s13059-016-0947-7.
115. Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095. https://doi.org/10.1038/nmeth.2645.
116. Andrews, T.S., and Hemberg, M. (2019). M3Drop: dropout-based feature selection for scRNASeq. Bioinformatics *35*, 2865–2867. https://doi.org/10.1093/bioinformatics/bty1044.
117. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
118. Burkhardt, D.B., Stanley, J.S., Tong, A., Perdigoto, A.L., Gigante, S.A., Herold, K.C., Wolf, G., Giraldez, A.J., van Dijk, D., and Krishnaswamy, S. (2021). Quantifying the effect of experimental perturbations at single-cell resolution. Nat. Biotechnol. *39*, 619–629. https://doi.org/10.1038/s41587-020-00803-5.
119. Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nat. Methods *13*, 845–848. https://doi.org/10.1038/nmeth.3971.
120. Hoffman, G.E., and Roussos, P. (2021). Dream: powerful differential expression analysis for repeated measures designs. Bioinformatics *37*, 192–201. https://doi.org/10.1093/bioinformatics/btaa687.
121. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.
122. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.
123. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.
124. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. *12*, 5692. https://doi.org/10.1038/s41467-021-25960-2.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
BV421 anti-human CD38 (clone HIT2)	BioLegend	Cat# 303526; RRID:AB_10983072
BV605 anti-human CD10 (clone HI10a)	BioLegend	Cat# 312222; RRID:AB_2562157
BV650 anti-human CD49f (clone GoH3)	BioLegend	Cat# 313629; RRID:AB_2686989
BV785 anti-human CD117 (clone 104D2)	BioLegend	Cat# 313238; RRID:AB_2629837
BB515 anti-human CD45RA (clone HI100)	BD	Cat# 564552; RRID:AB_2738841
PE anti-human CD123 (clone 6H6)	BioLegend	Cat# 306006; RRID:AB_314580
PE/Dazzle 594 anti-human CD49f (clone GoH3)	BioLegend	Cat# 313626; RRID:AB_2616782
PE/Cy7 anti-human CD90 (clone 5E10)	BioLegend	Cat# 328124; RRID:AB_2561693
APC anti-human CD34 (clone 581)	BioLegend	Cat# 343510; RRID:AB_1877153
PE/Dazzle 594 anti-human CD117 (clone 104D2)	BioLegend	Cat# 313226; RRID:AB_2566213
PE/Cy5 anti-human CD2 (clone RPA-2.10)	BioLegend	Cat# 300210; RRID:AB_314034
PE/Cy5 anti-human CD3 (clone HIT3a)	BioLegend	Cat# 300310; RRID:AB_314046
PE/Cy5 anti-human CD4 (clone RPA-T4)	BioLegend	Cat# 300510; RRID:AB_314078
PE/Cy5 anti-human CD8a (clone RPA-T8)	BioLegend	Cat# 301010; RRID:AB_314128
PE/Cy5 anti-human CD11b (clone ICRF44)	BioLegend	Cat# 301308; RRID:AB_314159
PE/Cy5 anti-human CD14 (clone 61D3)	eBioscience	Cat# 15-0149-42; RRID:AB_2573058
PE/Cy5 anti-human CD19 (clone HIB19)	BioLegend	Cat# 302210; RRID:AB_314240
PE/Cy5 anti-human CD20 (clone 2H7)	BioLegend	Cat# 302308; RRID:AB_314256
PE/Cy5 anti-human CD56 (clone MEM188)	BioLegend	Cat# 304608; RRID:AB_314450
PE/Cy5 anti-human CD235ab (clone HIR2)	BioLegend	Cat# 306606; RRID:AB_314623
FITC anti-human CD45RA (clone HI100)	BD	Cat# 555488; RRID:AB_395879
PE anti-human CD90 (clone 5E10)	BD	Cat# 555596; RRID:AB_395970
BV711 anti-human CD19 (clone SJ25C1)	BD	Cat# 563036; RRID:AB_2737968
PE/Cy5 anti-human CD49f (clone GoH3)	BD	Cat# 551129; RRID:AB_394062
APC anti-human CD271 (clone ME20.4-1.H4)	Miltenyi	Cat# 130-113-418; RRID:AB_2733363
APC/Cy7 anti-human CD34 (clone 581)	BD	Custom-made
PE/Cy7 anti-human CD38 (clone HB7)	BD	Cat# 335790; RRID:AB_399969
Alexa Fluor 700 anti-human CD10 (clone HI10a)	BD	Cat# 563509; RRID:AB_2738247
BV605 anti-human CD14 (clone M5E2)	BD	Cat# 564054; RRID:AB_2687593
V500 anti-human CD45 (clone HI30)	BD	Cat# 560777; RRID:AB_1937324
BV421 anti-human CD33 (clone WM53)	BioLegend	Cat# 303416; RRID:AB_2561690
Chemicals, peptides, and recombinant proteins		
RPMI-1640	Gibco	Cat# 21875034
IMDM	Gibco	Cat# 21056023
DPBS	ThermoFisher	Cat# 14190169
Fetal Bovine Serum	Sigma-Aldrich	Cat# F7524
Heparin sodium	Sigma-Aldrich	Cat# H3149
D-(+)-Glucose (dextrose)	Sigma-Aldrich	Cat# G7528
Citric Acid, trisodium salt, dihydrate	Sigma-Aldrich	Cat# C3434
Citric Acid, anhydrous	Sigma-Aldrich	Cat# C0759
Penicillin-Streptomycin (10,000 U/mL)	Gibco	Cat# 15140122
DNase I	Roche	Cat# 11284932001
Ficoll-Paque PLUS	GE Healthcare	Cat# 17-1440-03

(*Continued on next page*)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Propidium iodide	Biolegend	Cat# 421301
Propidium iodide	BD	Cat# 5564463
7-AAD	Biolegend	Cat# 420404
SDS 10%	Sigma-Aldrich	Cat# L4509
Triton X-100	Sigma-Aldrich	Cat# T8787
dNTPs (10 mM each)	Thermo Scientific	Cat# R0193
Poly-ethylene Glycol 8000 (40% solution)	Sigma-Aldrich	Cat# P1458
Recombinant RNAse Inhibitor (40 U/μL)	Takara	Cat# 2313B
Protease	Qiagen	Cat# 19155
Nuclease-free water	Invitrogen	Cat# AM9937
Tris-HCl 1M pH 8.0	Thermo Scientific	Cat# 15893661
NaCl 5M	Invitrogen	Cat# AM9760G
MgCl2 1M	Invitrogen	Cat# AM9530G
GTP Solution, Tris buffered	Thermo Scientific	Cat# R1461
Dithiothreitol (DTT), 0.1M Solution	Thermo Scientific	Cat# 707265ML
UltraPure Agarose	Invitrogen	Cat# 16500-500
Ethidium Bromide Solution	Invitrogen	Cat# 15585-011
MseI restriction enzyme	New England Biolabs	Cat# R0525S
ddPCR Supermix for Probes (No dUTP)	Bio-Rad	Cat# 186-3023
Critical commercial assays		
CD34 MicroBead kit	Miltenyi	Cat# 130-046-703
LS Columns	Miltenyi	Cat# 130-042-401
MycoAlert Mycoplasma Detection Kit	Lonza	Cat# LT07-218
CompBeads	BD Biosciences	Cat# 552843
DNeasy Blood & Tissue Kit	Qiagen	Cat# 69506
KAPA HyperPlus Kit	Roche	Cat# 07962428001
KAPA Dual-Indexed Adapter Set	Roche	Cat# 08278555702
HyperCap Target Enrichment Kit	Roche	Cat# 08286345001
HyperCap Bead Kit	Roche	Cat# 08286400001
SeqCap EZ Prime Choice Library (custom probes)	Roche	Cat# 08247480001
AMPure XP Beads	Beckman Coulter	Cat# A63881
Qubit dsDNA HS Assay Kit	Invitrogen	Cat# Q32854
ERCC RNA Spike-In Mix	Invitrogen	4456740
Maxima H Minus Reverse Transcriptase (200 U/μL)	Thermo Scientific	Cat# EP0753
KAPA HiFi HotStart ReadyMix	Roche	Cat# 07958935001
KAPA 2G Robust HS Ready Mix	Sigma-Aldrich	Cat# KK5702
FastStart High Fidelity PCR System, dNTPack	Sigma-Aldrich	Cat# 4738292001
Nextera XT DNA Library Preparation Kit	Illumina	Cat# FC-131-1096
Nextera XT Index Kit Set v2 Set A	Illumina	Cat# FC-131-2001
Nextera XT Index Kit Set v2 Set C	Illumina	Cat# FC-131-2003
Access Array Barcode Library for Illumina Sequencers-384, Single Direction	Fluidigm	Cat# 100-4876
High Sensitivity NGS Fragment Analysis Kit (1 bp–6,000 bp)	Agilent	Cat# DNF-474-0500
Agilent High Sensitivity DNA Kit	Agilent	Cat# 5067-4626
Agilent Tapestation HS D1000 ScreenTape	Agilent	Cat# 5067-5583
Agilent Tapestation HS D1000 Reagents	Agilent	Cat# 5067-5584

(*Continued on next page*)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Targeted DNA sequencing, raw data	This paper	EGA: EGAS00001007358
TARGET-seq+ single-cell RNA sequencing, raw data	This paper	EGA: EGAS00001007358
TARGET-seq+ single-cell genotyping, raw data	This paper	EGA: EGAS00001007358
TARGET-seq+ single-cell RNA sequencing, processed raw counts	This paper	Figshare: https://doi.org/10.25446/oxford.23576379
TARGET-seq+ single-cell genotyping data, processed allelic counts	This paper	Figshare: https://doi.org/10.25446/oxford.23576421
TARGET-seq+ single-cell metadata and genotypes	This paper	Figshare: https://doi.org/10.25446/oxford.23576262
In-house aged and young bone marrow snRNA-seq, raw data	This paper	GEO: GSE235646
Ainciburu aged and young bone marrow scRNA-seq, raw data	Ainciburu et al.[84]	GEO: GSE180298
GENCODE human gene annotation (v38)	GENCODE project	https://www.gencodegenes.org/human/release_38.html
cisTarget databases (Homo sapiens - hg38 - mc_v10_clust)	Aerts lab; Van de Sande et al.[101]	https://resources.aertslab.org/cistarget/databases/
Experimental models: Cell lines		
JURKAT	ATCC	RRID:CVCL_0367
Oligonucleotides		
OligodT-ISPCR (HPLC purification): AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTN	Picelli et al.[102]	N/A
Barcoded oligodT-ISPCR primers	Biomers (design: Rodriguez-Meira et al.[59])	N/A
TSO-LNA (RNase Free HPLC purification): AAGCAGTGGTATCAACGCAGAGTACATrGrG+G	IDT (design: Picelli et al.[102])	N/A
ISPCR primer (HPLC purification): AAGCAGTGGTATCAACGCAGAGT	IDT (design: Picelli et al.[102])	N/A
See Table S4 for target-specific genotyping primers used in the pre-amplification step (RNase Free HPLC purification)	IDT (design: this paper)	N/A
See Table S4 for target-specific nested barcoded genotyping primers used in the PCR1 barcoding step (Standard desalting)	IDT (design: this paper)	N/A
See Table S4 for custom transcriptome i5 index primers (HPLC purification)	IDT (design: this paper)	N/A
See Table S4 for the list of probes used for ddPCR (HPLC purification)	Biomers (design: this paper)	N/A
See Table S4 for the list of primers used for ddPCR (Standard desalting)	IDT (design: this paper)	N/A
P5-SEQ primer (PAGE purification): GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTTGC*T	Rodriguez-Meira et al.[59]	N/A
I5-SEQ primer (PAGE purification): AGCAACTCTGCGTTGATACCACTGCTTCCGCGGACAGG*C	IDT (design: this paper)	N/A

(*Continued on next page*)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LCS1 sequencing primer (HPLC purified): GGCGACCACCGAGATCTACACTGACG ACATGGTTCTACA	IDT	N/A
CS2 sequencing primer (HPLC purified): T+AC+GGT+AGCAGAGACTTGGTCT	IDT	N/A
CS2rc sequencing primer (HPLC purified): A+GAC+CA+AGTCTCTGCTACCGTA	IDT	N/A
Software and algorithms		
Bcl2fastq (v2.20)	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
Python (v3)	Python Software Foundation	https://www.python.org
CGAT-core	Bioconda	https://github.com/cgat-developers/cgat-core
Burrows-Wheeler aligner (v0.7.17)	Li and Durbin[103]	https://bio-bwa.sourceforge.net/
GATK (v4.1.2.0)	Broad Institute	https://software.broadinstitute.org/gatk/
Picard (v2.21.3)	Broad Institute	https://broadinstitute.github.io/picard/
VarDictJava (v1.7.0)	Lai et al.[104]	https://github.com/AstraZeneca-NGS/VarDictJava
Bcftools (v1.9)	HTSlib	https://samtools.github.io/bcftools/
Annovar (v20191024)	Yang et al.[105]	https://annovar.openbioinformatics.org/
Integrative Genomics Viewer (IGV; v2.14.1)	Broad Institute	https://software.broadinstitute.org/software/igv/
Primer3Plus	Untergasser et al.[106]	https://www.primer3plus.com
Primer-BLAST	National Library of Medicine	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
TARGET-seq genotyping pipeline	Github; Rodriguez-Meira et al.[59]	https://github.com/albarmeira/TARGET-seq
infSCITE	Github	https://github.com/cbg-ethz/infSCITE
Samtools	HTSlib	http://www.htslib.org/download/
Cutadapt (v3.4)	Bioconda	https://cutadapt.readthedocs.io/en/stable/
FastQC (v0.11.9)	Bioconda	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC (v1.11)	Bioconda	https://multiqc.info
STAR (v2.7.10a)	Github	https://github.com/alexdobin/STAR
R (v4.2.1)	R-Project	https://www.r-project.org
ggplot2 (v3.3.6)	CRAN	https://ggplot2.tidyverse.org
SingCellaR (v1.2.1)	Github	https://github.com/supatt-lab/SingCellaR
Scran (v1.24.0)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/scran.html
M3Drop (v1.22.0)	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/M3Drop.html
Harmony (v0.1.0)	CRAN	https://github.com/immunogenomics/harmony
variancePartition (v1.26.0)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/variancePartition.html
fgsea (v1.22.0)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/fgsea.html
AUCell (v1.18.1)	Bioconductor	https://bioconductor.org/packages/release/bioc/html/AUCell.html
Scanpy (v1.9.1)	Conda-forge	https://scanpy.readthedocs.io/en/stable/
MELD (v1.0.0)	PyPI	https://github.com/KrishnaswamyLab/MELD
pySCENIC (v0.12.0)	PyPI	https://github.com/aertslab/pySCENIC
FlowJo (v10.8.1)	BD	https://www.bdbiosciences.com/en-gb/products/software/flowjo-v10-software
BioRender	BioRender	https://biorender.com
Custom code for data analysis	This study	N/A

(*Continued on next page*)

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
12.5 mL GRIPTIP, Sterile, Filter	INTEGRA Biosciences	Cat# 6455
High Volume MANTIS Chip	FORMULATRIX	Cat# MCHVSMR6

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Paresh Vyas (paresh.vyas@imm.ox.ac.uk).

Materials availability

The list of all oligo sequences designed in this study and used for single-cell genotyping can be found in Table S4. These include both target-specific oligos used in the PCR after reverse transcription, and nested barcoded target-specific oligos used in genotyping PCR1. Barcoded oligodT-ISPCR primers were kindly provided by Prof. Adam Mead and Dr. Alba Rodriguez-Meira, and the sequences are listed in Table S4.

Data and code availability

- Raw targeted DNA sequencing data, TARGET-seq+ scRNA-seq, and TARGET-seq+ single-cell genotyping data have been deposited at European Genome-Phenome Archive (EGA: EGAS00001007358) in order to comply with ethical approvals and are available as of the date of publication. Processed TARGET-seq+ scRNA-seq, single-cell genotyping, and metadata are available through Figshare. Single-nucleus RNA-seq data for the in-house aged and young bone marrow dataset have been deposited in GEO (GEO: GSE235646). Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

JURKAT human cell line was cultured in RPMI-1640 medium (Cat# 21875034, Gibco) supplemented with 10% FBS and 1% V/V Pen-Strep (Cat# 15140122, Gibco). Cells were regularly screened for Mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Cat# LT07-218). Cells were passaged every 2-3 days and seeded at approximately 500,000 cells/mL. Cell lines were kept in a CO_2 incubator at 37°C.

Patient samples

Patient samples were collected from individuals undergoing elective total hip replacement (THR) surgery at the Nuffield Orthopaedic Centre, Oxford, under the Mechanisms of Age-Related Clonal Haematopoiesis (MARCH) Study. Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. This study was approved by the Yorkshire & The Humber - Bradford Leeds Research Ethics Committee (NHS REC Ref: 17/YH/0382). Exclusion criteria were: History of rheumatoid arthritis or other inflammatory arthritis, history of septic arthritis in the limb undergoing surgery, history of hematological cancer, bisphosphonate use, and oral steroid use. Patient characteristics are summarized in Table S1.

For the multi-ome analysis in young versus aged human bone marrow, bone marrow cells from young donors (26-year-old female and 24-year-old male) were purchased from Lonza, while bone marrow samples from aged donors (70 and 77-year-old females) undergoing hip replacement surgery were collected at the Traumatology and Orthopedics Hospital Lomas Verdes (IMSS), Mexico. These elderly donors were confirmed to have no dysplasia of any hematopoietic lineages by histological and CBC analysis.[107] Ethical approval was obtained from the Institutional Review Board (R-2012-785-092). Patient consent was obtained verbally, and as determined by the Institutional Ethical Board.

METHOD DETAILS

Sample collection and processing

Trabecular bone fragments and bone marrow aspirates were obtained from the femoral canal and collected in 10 mL anticoagulated buffer containing acid-citrate-dextrose, heparin sodium and DNase. Samples of peripheral blood were collected in EDTA vacutainers.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

All samples were processed within 24 hours of collection. Peripheral blood and bone marrow aspirate samples were diluted 1:1 in RPMI-1640 (Gibco) and filtered through a 70 μm cell strainer. Trabecular bone samples were manually fragmented with scissors and washed thoroughly in RPMI media with DNase to collect trabecular marrow, which was then filtered through a 70 μm cell strainer to obtain a single cell suspension and combined with the bone marrow aspirate. Mononuclear cells were then isolated by Ficoll density gradient separation (Sigma-Aldrich). For some samples, bone marrow $CD34^+$ cells were purified using a CD34 MicroBead kit and MACS separation columns (Miltenyi Biotec), according to the manufacturer's instructions. Unseparated MNCs, CD34-enriched and CD34-depleted fractions were frozen in 90% fetal bovine serum (FBS, Sigma-Aldrich) with 10% dimethyl sulfoxide (DMSO) and stored in liquid nitrogen until further use.

Peripheral blood granulocytic cell pellets isolated by Ficoll density gradient centrifugation were frozen for DNA sequencing analysis of mutations in peripheral blood.

Targeted DNA sequencing

Library preparation and sequencing

Targeted DNA sequencing was performed on bone marrow MNCs and peripheral blood granulocytic DNA samples. Pre-capture DNA libraries were prepared using the KAPA HyperPlus protocol (Roche). 100 ng of genomic DNA was fragmented by enzymatic fragmentation. Following end repair and A-tailing, adapter ligation was performed using KAPA dual-indexed adapters (Roche). Library cleanup and double-sided size selection was performed using Agencourt AMPure XP beads (Beckman Coulter) to obtain fragments of ~320 bp. Libraries were amplified by ligation-mediated PCR for 6 cycles using a KAPA HiFi HotStart high-fidelity DNA polymerase (Roche) and purified using AMPure XP beads.

Targeted capture was performed using a custom pool of biotinylated capture probes (SeqCap EZ Prime Choice, Roche) targeting 97 genes recurrently mutated in myeloid malignancies and clonal hematopoiesis spanning 347 kb (Table S2). Amplified DNA libraries were hybridized to the capture probes in pools of 10-12 samples according to the manufacturer's instructions. The captured DNA libraries were amplified by 14 cycles of PCR using a KAPA HiFi HotStart DNA polymerase and purified using AMPure XP Beads.

Post-capture amplified DNA libraries were quantified by Qubit (Life Technologies) and size distribution and quality analyzed using a Bioanalyzer chip (Agilent Technologies). Libraries were pooled in equimolar concentrations and were sequenced on an Illumina NextSeq 500 using paired-end reads.

Somatic variant calling

Sequencing data were analyzed with a custom Python pipeline based on the GATK Best Practices (GATK v4.1.2.0 and Picard v2.21.3). Raw sequencing reads were converted to an unmapped BAM file and adapter sequences soft-clipped using Picard *MarkIlluminaAdapters*. Following conversion back to a FASTQ file, reads were mapped to the hg38 human reference genome assembly using the Burrows–Wheeler aligner v0.7.17[103] with alternate contig-aware alignment. Mapped and unmapped BAM files were merged using *MergeBamAlignment* and reads from different sequencing lanes were combined. Duplicate reads were marked using Picard *MarkDuplicates* and base quality scores recalibrated with GATK BaseRecalibrator and ApplyBQSR. Somatic variant calling was performed on the pre-processed BAM files using VarDictJava v1.7.0[104] and Mutect2[108] in tumor-only mode. For VarDict, variants were called with a minimum variant allele frequency of 0.01, minimum base quality score of 25 and minimum supporting reads of 2, with indel realignment and removal of adapter sequences. For Mutect2, a minimum tumor LOD of 2 was used and variants were filtered for sequence context-dependent artefacts using *FilterMutectCalls* and *FilterByOrientationBias*. Indels were left-aligned and normalized using bcftools norm (v1.9). Variants were annotated using Annovar.[105] Target enrichment metrics and coverage was calculated using Picard *CollectHsMetrics* and custom scripts.

VarDict and Mutect2 variant calls were analyzed separately to identify a consensus list of high-confidence variants. The following post-processing filters were applied to VarDict calls to exclude likely sequencing artefacts:

1. Minimum of 5 variant reads for SNVs (with at least 2 reads in forward and reverse directions), or minimum of 10 variant reads for indels (with at least 4 reads in forward and reverse directions).
2. Minimum base quality score 30.
3. Minimum mapping quality score 40, except for variants in *U2AF1;U2AF1L5*, where the mapping quality was ignored. This is because in hg38, there is a duplication of the *U2AF1* gene on chromosome 21 called *U2AF1L5*, which results in reads being flagged as multi-mapped.
4. Maximum strand bias Fisher *p*-value of 0.0001.
5. No position bias towards beginning or end of reads.

The following post-processing filters were applied to Mutect2 calls:

1. Passed all default Mutect2 filters or only failing the *clustered_events* filter.
2. Minimum of 5 variant reads for SNVs (with at least 2 reads in forward and reverse directions), or minimum of 10 variant reads for indels (with at least 4 reads in forward and reverse directions).

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Variants were flagged as likely germline, or sequencing artefacts, if any of the following applied:

1. Variant allele with a population allele frequency >1 in 1,000 according to any of three large polymorphism databases (Gnomad, 1000 Genomes Project, ESP6500) that is not a hotspot driver mutation with a COSMIC (v88) occurrence count of >100 cases or is present in a list of clonal hematopoiesis-associated mutations compiled from five large studies.[14,15,17,19,109]
2. Variant allele frequency (VAF) between 0.4–0.6 or >0.9 unless recurrent in COSMIC >5 times, or previously reported in clonal hematopoiesis.
3. Present in a panel of normal cord blood samples.
4. Recurrent in the cohort unless present at least 5 times in COSMIC or at least 2 times in the clonal hematopoiesis studies.

After filtering, variants were manually inspected using the Integrative Genomics Viewer (IGV) tool (http://software.broadinstitute.org/software/igv/).

Annotation of pathogenic driver variants

Samples were annotated as having CH based on the presence of at least one driver mutation in bone marrow sequencing at VAF $\geq$ 0.01. Variants were annotated as pathogenic driver mutations using the following criteria:

1. Truncating mutations (nonsense, splice site or frameshift indel) in the following genes: *DNMT3A*, *TET2*, *ASXL1*, *NF1*, *IKZF1*, *RAD21*, *WT1*, *KMT2D*, *SH2B3*, *TP53*, *CEBPA*, *RUNX1*, *BCOR*, *KDM6A*, *STAG2*, *PHF6*, *KMT2C*, *KMT2E*, *PPM1D*, *ATRX*, *EZH2*, *CREBBP*, *NOTCH1*, *CUX1* and *ZRSR2*.
2. Non-synonymous variants at the following hotspot residues: *CBL* E366, L380, C384, C396, C404 and R420; *DNMT3A* R882; *FLT3* D835; *IDH1* R132; *IDH2* R140 and R172; *JAK2* V617F; *KIT* W557, V559 and D816; *KRAS* G12, G13, Q61 and A146; *MPL* W515; *NRAS* G12, G13 and Q61; *SF3B1* K666 and K700; *SRSF2* P95; *U2AF1* S34, R156 and Q157.
3. Non-synonymous variants occurring within the following residues of *DNMT3A*: p.292-350, p.482-614 and p.634-912; *TET2*: p.1104-1481 and p.1843-2002; or *NOTCH1*: p.1574-1620, p.1671-1721.
4. Truncating variants in *CALR* exon 9.
5. *FLT3* internal tandem duplications.
6. Non-synonymous variants reported at least 10 times in COSMIC with VAF < 0.4.
7. Non-synonymous variants falling within an annotated InterPro domain with VAF < 0.4.
8. Non-synonymous variants reported in COSMIC > 100 times.

If a variant did not meet these criteria, it was annotated as a variant of unknown significance (VUS).

Comparison of bone marrow and peripheral blood allele frequencies

Variants were called in peripheral blood DNA sequencing data as described above, using a minimum VAF cutoff of 0.005. Unfiltered variant calls were intersected with the list of curated bone marrow variants to compare the VAF between bone marrow and blood. For those mutations known to be present in bone marrow which were not called in peripheral blood, raw allele counts were performed directly from the BAM files using bcftools mpileup (with minimum base quality of 30 and minimum mapping quality of 35), and the allele frequency was calculated.

Flow cytometry and FACS sorting

Thawing media was prepared with IMDM medium (Gibco) supplemented with 20% FBS and 110 μg/mL DNase. Bone marrow samples were thawed at 37°C in a water bath, 1 mL warm FBS was added, and the suspension then diluted by dropwise addition of 8 mL thawing media. The suspension was centrifuged at 400 × g for 10 mins, cells were resuspended in flow cytometry staining medium (IMDM with 10% FBS and 10 μg/mL DNase), filtered through a 35 μm cell strainer, and placed on ice.

Cells were stained with antibodies listed for 20–30 min on ice. Following antibody incubations, cells were washed with 1 mL flow cytometry staining buffer, centrifuged at 350 × g for 5 min, and resuspended in flow cytometry staining buffer containing the live/dead stain.

For flow cytometry analysis of BM samples, unseparated BM MNCs were used. Samples were stained with the following antibodies: anti-CD38-BV421 (1:20, Biolegend, clone HIT2), anti-CD10-BV605 (1:40, Biolegend, clone HI10a), anti-CD49f-BV650 (1:40, Biolegend, clone GoH3), anti-CD117-BV785 (1:40, Biolegend, clone 104D2), anti-CD45RA-BB515 (1:40, BD, clone HI100), anti-CD123-PE (1:40, Biolegend, clone 6H6), anti-CD90-PE/Cy7 (1:20, Biolegend, clone HI100), anti-CD34-APC (1:160, Biolegend, clone 581), anti-CD2-PE/Cy5 (1:160, Biolegend, clone RPA-2.10), anti-CD3-PE/Cy5 (1:320, Biolegend, clone HIT3a), anti-CD4-PE/Cy5 (1:160, Biolegend, clone RPA-T4), anti-CD8a-PE/Cy5 (1:320, Biolegend, clone RPA-T8), anti-CD11b-PE/Cy5 (1:160, Biolegend, clone ICRF44), anti-CD14-PE/Cy5 (1:160, eBioscience, clone 61D3), anti-CD19-PE/Cy5 (1:160, Biolegend, clone HIB19), anti-CD20-PE/Cy5 (1:160, Biolegend, clone 2H7), anti-CD56-PE/Cy5 (1:80, Biolegend, clone MEM188), anti-CD235ab-PE/Cy5 (1:320, Biolegend, clone HIR2), and propidium iodide (final concentration 3 μM; Biolegend) was used for dead cell exclusion. Analysis was done on a LSR Fortessa X20 (BD Biosciences). Unstained, single-stained (CompBeads, BD Biosciences), and fluorescence-minus-one (FMO) controls were used to determine background staining and compensation in each channel. Gating was kept consistent across all samples to enable quantification of population sizes.

For FACS sorting of BM samples for TARGET-seq+ and ddPCR, either unseparated BM MNCs or CD34-enriched BM MNCs were used. Samples were stained with the following antibodies: anti-CD38-BV421 (1:20, Biolegend, clone HIT2), anti-CD10-BV605 (1:40,

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Biolegend, clone HI10a), anti-CD117-BV785 (1:40, Biolegend, clone 104D2), anti-CD45RA-BB515 (1:40, BD, clone HI100), anti-CD123-PE (1:40, Biolegend, clone 6H6), anti-CD49f-PE/Dazzle594 (1:160, Biolegend, clone GoH3), anti-CD90-PE/Cy7 (1:20, Biolegend, clone HI100), anti-CD34-APC (1:160, Biolegend, clone 581), anti-CD2-PE/Cy5 (1:160, Biolegend, clone RPA-2.10), anti-CD3-PE/Cy5 (1:320, Biolegend, clone HIT3a), anti-CD4-PE/Cy5 (1:160, Biolegend, clone RPA-T4), anti-CD8a-PE/Cy5 (1:320, Biolegend, clone RPA-T8), anti-CD11b-PE/Cy5 (1:160, Biolegend, clone ICRF44), anti-CD14-PE/Cy5 (1:160, eBioscience, clone 61D3), anti-CD19-PE/Cy5 (1:160, Biolegend, clone HIB19), anti-CD20-PE/Cy5 (1:160, Biolegend, clone 2H7), anti-CD56-PE/Cy5 (1:80, Biolegend, clone MEM188), anti-CD235ab-PE/Cy5 (1:320, Biolegend, clone HIR2), and 7-AAD (Biolegend) was used for dead cell exclusion. For sample NOC156, cells were stained with the same panel except anti-CD49f, and anti-CD117 was substituted for anti-CD117- PE/Dazzle594 (1:80, Biolegend, clone 104D2). Single-cell index sorting was performed on a Sony MA900 into 384-well plates containing 3 μL lysis buffer (except for optimization experiments, which were done in 96-well plates). Unstained, single-stained, and FMO controls were used to determine background staining and compensation in each channel. Doublets and dead cells were excluded. The following populations were sorted: Live/Lin$^-$CD34$^+$, Live/Lin$^-$ CD34$^+$CD38$^-$, and CD34$^-$ cells (except for sample NOC156, where only CD34$^+$ cells were analyzed). In addition to the sample of interest, cells from the NOC153 control sample were sorted onto every plate, making up approximately 10% of wells on each plate, and two empty wells were used as no-template controls. After sorting, each plate was centrifuged and snap frozen on dry ice prior to storage at –80°C.

Flow cytometry data analysis was performed using FlowJo v10.8.1 and R.

Whole-genome sequencing (WGS) analysis

For validation of clonal cell fractions detected by TARGET-seq+, we analyzed data from our separate study in which WGS was performed on FACS-sorted Lin$^-$CD34$^+$ cells.[110] bcftools mpileup was used to calculate the VAF of each mutation from the BAM files, and the mutant cell fraction was calculated as 2 × VAF for autosomal mutations.

Droplet digital PCR (ddPCR)

For CH samples NOC002 (2 loci), NOC115 (2 loci), and NOC132 (1 locus), ddPCR on sorted immunophenotypic populations was performed to validate the frequency of mutations screened by TARGET-seq+. For each locus, we first determined the optimal DNA strand for the WT probe to outcompete the binding of the mutant probe to the WT allele, and vice-versa, using the T_m mismatch option from the IDT OligoAnalyzer tool. Next, for each locus, primers and probes were designed with Primer3Plus, using the Internal Oligo option. Binding specificity of the primers and of the reference allele probes was checked with Primer BLAST, and the Hetero-Dimer option from the IDT OligoAnalyzer was used to control for primer and probe cross-binding. For all loci, the WT probe contained the HEX fluorophore, and the mutant probe contained the 6-FAM fluorophore at the 5' end.

Primers for all loci were first tested for specificity and successful amplification in a bulk PCR. Next, the optimal annealing temperature for each primer/probe set was determined by an annealing temperature gradient experiment, using gDNA from the respective samples (extracted either from BM MNCs or PB granulocytes) for the mutant allele probes, or a WT control gDNA (extracted from PB of a WT donor) for the reference allele probes.

BM samples were then thawed and stained with antibodies for FACS sorting, as detailed above. Up to 1000 cells from immunophenotypic populations of interest were sorted on purity mode into tubes containing 5 μL of lysis buffer consisting of 40 mM Tris-HCl pH 8.0 (Thermo Scientific), 0.2% Triton X-100 (Sigma-Aldrich), and 2 x 10^{-04} AU/μL protease (Qiagen). Where possible, multiple replicates (2-3) of the same population were sorted. Cell pellets were incubated at 56°C for 1 h, followed by incubation at 72°C for 15 min to perform lysis and protease heat-inactivation. PCR was set up using 11 μL of 2X ddPCR supermix for probes (no dUTP), 1.1 μL of the 20X primer/probe mix (containing each primer at 18 μM and each probe at 5 μM), 0.22 μL of MseI enzyme, 3-5 μL of cell lysate or gDNA, and 4.7-6.7 μL of nuclease-free water. Droplets were generated using a Bio-Rad Automated Droplet Generator. The following PCR program was used: 95°C for 10 min for initial denaturation, 40 cycles of 94°C for 30 s and 59/60/63°C for 1 min. Final incubation was performed at 98°C for 10 min. Droplets were read with a Bio-Rad QX200 Droplet Reader. No-template controls, WT gDNA negative controls, and sample gDNA positive controls were used to set gates for mutant, WT, and double positive droplets. Fractional abundance (VAF) in each replicate was calculated with the QX Manager Software v2.1.

TARGET-seq+ library preparation

Primer design

Targeted genotyping primers used in the pre-amplification RT-PCR step were designed to amplify regions 180–900 bp long. Where possible, gDNA primers were designed to anneal within intronic regions flanking the mutation of interest, while mRNA primers were designed to anneal to exonic regions outside of the gDNA amplicon, so that independent amplicons would be generated from mRNA and gDNA. Furthermore, when a heterozygous SNP was observed close to the mutation, primers were placed in order to cover the SNP within the amplicon, enabling a direct measurement of allelic dropout. Primers were designed with Primer3Plus[106] and specificity was checked using Primer-BLAST.[111] For each target, primer pairs were tested for specificity and efficiency in bulk PCR reactions and in single cells.

Targeted primers for use in the genotyping PCR1 step were designed to be nested within each of the amplicons generated in the pre-amplification RT-PCR. Nested amplicons were 290–631 bp in length. As for pre-amplification primers, gDNA primers were designed to anneal within intronic regions flanking the mutation, and cDNA primers were designed to anneal to exonic regions outside of

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

the gDNA amplicon, to generate independent mutational readouts from cDNA and gDNA, where possible. Primer pairs were tested for specificity and efficiency in bulk PCR reactions and in single cells.

Lysis buffer preparation

Lysis buffer was prepared as described in Table S4, consisting of 0.1% Triton X-100 (Sigma-Aldrich), 0.5 mM dNTPs (Life Technologies), 5% PEG 8000 (Sigma-Aldrich), 0.5 U/μL RNase inhibitor (Takara), 2.7 x 10^{-05} AU/μL protease (Qiagen), and 1:8,000,000 diluted ERCC spike-in mix (Ambion). 25 μL of lysis buffer was dispensed into each well of a 384-well stock plate using a Formulatrix Mantis with a high-volume chip, and 1.79 μL of 10 μM barcoded oligodT-ISPCR primer was added to each well using an INTEGRA Viaflow. 3 μL of barcoded oligodT-lysis buffer mix was then transferred into each well of a 384-well plate. Plates were sealed and stored at –80°C and thawed prior to cell sorting.

Reverse transcription and pre-amplification

Plates containing sorted cells were removed from –80°C storage and incubated at 72°C for 15 min to perform cell lysis, RNA denaturation, and protease heat inactivation. For reverse transcription (RT), 1 μL of RT mix was added, bringing reaction concentrations to 25 mM Tris-HCl (Thermo Scientific), 30 mM NaCl (Invitrogen), 2.5 mM $MgCl_2$ (Invitrogen), 1 mM GTP (Thermo Scientific), 8 mM Dithiothreitol (DTT, Thermo Scientific), 0.5 U/μl RNase inhibitor (Takara), 2 μM of Smart-seq2 template switching oligo (TSO, IDT), 2 U/μl of Maxima H-minus reverse transcriptase enzyme (Thermo Scientific), and target-specific mRNA primers (70 nM final concentration). RT was performed by incubation at 42°C for 90 min followed by 10 cycles of 50°C for 2 min and 42°C for 2 min. The reaction was terminated by incubating at 85°C for 5 min.

Pre-amplification PCR mix containing target-specific genotyping primers was prepared as described in Table S4, to achieve reaction concentrations of 1× KAPA HiFi HotStart Ready Mix (Roche), 50 nM ISPCR primer, 28 nM target-specific cDNA primers, and 400 nM target-specific gDNA primers. Pre-amplification PCR was performed directly after reverse transcription by addition of 6 μL PCR mix and incubation on a thermocycler using the following program: 98°C for 3 min for initial denaturation, 21 cycles of 98°C for 20 s, 67°C for 30 s and 72°C for 6 min. Final elongation was performed at 72°C for 5 min. Conditions used for all RT-PCR steps are listed in Table S4. The sequences of the primers used in the RT-PCR steps for whole transcriptome amplification and targeted genotyping amplification are listed in Table S4.

Following cDNA amplification, successful libraries contain whole transcriptome cDNA and amplicons spanning each targeted mutation. An aliquot of this cDNA-amplicon mix was used for whole transcriptome library preparation and another aliquot for single-cell genotyping library preparation. 1 μL of cDNA-amplicon mix was pooled per well to create cDNA pools from 192 uniquely barcoded single-cell libraries, using a Mosquito HTS liquid handling platform (TTP Labtech). Each cDNA pool was purified twice using Ampure XP beads with 0.6:1 beads to cDNA ratio. Pooled cDNA libraries were checked using a High Sensitivity DNA Kit on a Bioanalyzer (Agilent) or a High Sensitivity NGS Fragment Analysis Kit (1 bp - 6,000 bp) on a Fragment Analyzer (Agilent). Libraries were quantified by Qubit dsDNA HS Assay (Life Technologies). These pools were used to generate 3' biased whole-transcriptome libraries. The remainder of the cDNA-amplicon mix was diluted 1:2 with water and stored at –20°C for use in single-cell genotyping.

Whole-transcriptome library preparation and sequencing

Bead-purified cDNA pools were used for tagmentation-based library preparation with a Nextera XT DNA Library Preparation Kit (Illumina) using a custom PCR amplification strategy to generate 3′ biased libraries containing oligodT cell barcodes as previously published,[59] with some modifications. Pooled cDNA libraries were diluted to 800 pg/μL and a total of 4 ng (5 μL) from each pool was used in the tagmentation reaction with 10 μL tagmentation buffer (TD) and 5 μL ATM enzyme. The reaction was incubated at 55°C for 10 min, followed by the addition of 5 μL 0.2% SDS to release Tn5 from the DNA. Library amplification was performed using 5 μL Nextera XT i7 forward index primer (Illumina) and 5 μL custom i5 index primers (2 μM) (see Table S4 for sequences). The custom i5 index primer binds the barcoded oligodT-ISPCR adapter, resulting in amplification of the 3' fragments containing the cell barcode. PCR was performed by adding NPM enzyme (Nextera XT DNA Library Preparation Kit, Illumina) and incubation on a thermocycler using the following program: 95°C for 30 s, 14 cycles of 95°C for 10 s, 55°C for 30 s and 72°C for 30 s, and then a final elongation of 5 min at 72°C. After tagmentation, each indexed pool was purified twice with Ampure XP beads using a 0.7:1 beads to cDNA ratio. Library quality was checked using a High Sensitivity DNA Kit on a Bioanalyzer and quantified using Qubit dsDNA HS Assay. Equimolar pools were made and sequenced using custom sequencing primers for Read1 and Index2 (P5-SEQ, I5-SEQ, 300 nM in HT1 buffer, see Table S4). For benchmarking experiments, libraries were sequenced on a NextSeq 500/550 High Output v2.5 (75 cycle) kit (Illumina) using the following sequencing configuration: 15 bp R1; 8 bp index read; 69 bp R2. For the main experiments, up to 9,984 single-cell libraries (52 pools of 192 single-cell libraries) were sequenced on a NovaSeq S4 flow cell with a targeted sequencing depth of 1 million reads/cell using the following sequencing configuration: 15 bp R1; 8 bp index read 1; 8 bp index read 2; 200 bp R2.

Targeted single-cell genotyping

To generate Illumina-compatible libraries for single-cell genotyping, two PCR steps were performed as previously published in the TARGET-seq protocol.[59] As the genotyping amplicons generated by the pre-amplification RT-PCR are not barcoded, genotyping PCR reactions were carried out separately for each single-cell library.

In the first PCR step (genotyping PCR1), nested target-specific primers containing universal CS1 (forward primer) or CS2 (reverse primer) adapters are used to amplify the target regions of interest. Incorporation of a barcode sequence specific to each plate into these primers enables libraries from different plates to be pooled subsequently. Primer sequences used for genotyping PCR1 for each sample are listed in Table S4. PCR1 reactions were performed using 3.25 μL of KAPA 2G Robust HS Ready Mix (Sigma-Aldrich), 1.5 μL of diluted cDNA-amplicon mix and 300 nM target-specific primers, in a 6.5 μL reaction.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

In the second PCR step (genotyping PCR2), Illumina-compatible adapters containing a 10 bp cell barcode are attached to the genotyping PCR1 product by binding to the CS1/CS2 adapters. PCR2 reactions were performed using FastStart High Fidelity polymerase (Sigma-Aldrich) with 1.0 μL of PCR1 product and 1.2 μL of each barcode primer mix (Access Array Barcode Library for Illumina Sequencers- 384, Single Direction, Fluidigm) in a 6.2 μL reaction.

Indexed amplicons were pooled using a Mosquito HTS liquid handling platform and purified with Ampure XP beads using a 0.8:1 beads to PCR product ratio. Purified pools were quantified using Qubit dsDNA HS Assay and the quality checked using a Tapestation High Sensitivity D1000 kit (Agilent) to ensure the size distribution of amplicons was as expected. Each pool was diluted to a final concentration of 4 nM and further diluted to 10 pM in HT1 buffer prior sequencing. Libraries were sequenced on a NextSeq 500/550 Mid Output v2.5 kit (300 cycle) (Illumina) using 150 bp paired-end reads, with 10 bp for the cell barcode index read and custom sequencing primers (Table S4).

TARGET-seq+ validation experiments

Validation experiments comparing TARGET-seq+ with TARGET-seq were performed in 96-well plates using JURKAT cells and primary human $CD34^+$ HSPCs. 3' TARGET-seq libraries were generated according to the published protocol.[59] Cells were sorted into 4.1 μL lysis buffer, consisting of 0.18% Triton X-100 (Sigma- Aldrich), 1.0 mM dNTP (Life Technologies), 1.0 U/μl RNase inhibitor (Takara), 2.7×10^{-5} AU/mL protease (Qiagen), 1.0 μM barcoded oligodT-ISPCR primer. RT was performed using SMARTScribe enzyme, RNase inhibitor, Smart-seq2 TSO (1 μM final concentration) and targeted mRNA primers (700 nM final concentration). PCR pre-amplification was performed using SeqAmp DNA polymerase, ISPCR primers (50 nM final concentration) and targeted gDNA and cDNA primers. TARGET-seq+ RT-PCR was performed as described above using double volumes per well. For both conditions, 20 cycles of amplification were used for JURKAT cells and 24 cycles of amplification for HSPCs. 3' transcriptome libraries were prepared as for TARGET-seq+ libraries detailed above and were sequenced on a NextSeq 500/550 High Output v2.5 (75 cycle) kit.

Targeted single-cell genotyping analysis

Pre-processing and mutation calling

Single-cell genotyping reads were pre-processed using the custom TARGET-seq pipeline (https://github.com/albarmeira/TARGET-seq).[59] Reads were first demultiplexed using the 384 well barcodes introduced via the genotyping PCR2 reaction, followed by demultiplexing based on plate barcodes introduced during genotyping PCR1. This generated separate FASTQ files for each single cell. Reads were aligned to hg38 using STAR version 2.7.3a with default settings and cDNA/gDNA amplicons were separated into different bam files, extracting reads matching the primer sequences used for targeted PCR barcoding. This allowed independent mutational information to be obtained from cDNA and gDNA amplicons. Variant calling was performed with *mpileup* (samtools version 1.1, options: *–minBQ 30*, *–count-orphans*, *–ignore overlaps*) and results were summarized using the custom pipeline.

For indels, reads from cDNA/gDNA amplicons were separated into different FASTQ files using cutadapt to extract reads matching the primer sequences. The number of reads with the wild-type or mutant sequence was counted for each cell using fastq-grep.

Mutational calling in single cells was then performed with custom R scripts, separately for each mutation. Coverage for each cell was calculated as the sum of all reads across the variant locus for that cell. Empty wells routinely displayed zero or very few reads (usually up to 2), indicating no cross-well contamination. A filtering threshold was applied to remove cells where the amplicon was not detected, or where coverage was too low for reliable genotyping. The minimum coverage was 50 reads for gDNA amplicons and 30 reads for cDNA amplicons. In cells with coverage below the threshold, the amplicon was called undetected.

The single-cell variant allele frequency (scVAF) for each cell was calculated with the following formula:

$$scVAF = variant\ reads \ / \ coverage$$

In a single cell, there are only two genomic DNA alleles. In the first cycles of PCR, if the amplification of one allele is favored over the other, there will be unequal amplification of the two alleles. Therefore, it is universal in single-cell genotyping to observe a distribution of scVAFs that will range from ~1% (0.01) to ~99% (0.99).[59,99,112] This scVAF distribution is also observed for genotyping of mRNA/cDNA. Furthermore, there is an inherent error rate of the DNA polymerase in PCR amplification and in next-generation sequencing, which needs to be accounted for when calling cell genotypes. To define the appropriate scVAF thresholds for CH mutation calling, we adopted a method used in our two prior studies.[99,112] Single cells from a control bone marrow sample that was WT for all mutations of interest were FACS-sorted onto every plate and processed in parallel with other wells containing single cells from a CH sample. Since the genotype of these control cells was known *a priori*, these data were used to determine the 'noise' or error rate at each locus. We thus used the scVAF distribution from these WT control cells ($scVAF_{WT\text{-}CTRL}$) to set thresholds for assigning a genotype to each single cell from samples of interest (Figure S3A). We define two thresholds for calling CH mutations. The lower scVAF threshold for calling a cell WT was set using the following formula:

$$scVAF\ threshold\ for\ calling\ a\ cell\ WT = mean(scVAF_{WT\text{-}CTRL}) + 3 \times SD(scVAF_{WT\text{-}CTRL})$$

Cells with scVAF below this threshold were called WT for that mutation. The scVAF threshold for calling a cell mutant was set using the following formula:

$$scVAF\ threshold\ for\ calling\ a\ cell\ mutant = mean(scVAF_{WT\text{-}CTRL}) + 3 \times SD(scVAF_{WT\text{-}CTRL}) + 0.01$$

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

Cells with scVAF above this threshold were called mutant for that mutation. For example, for the *TET2* pR1261C gDNA amplicon, the scVAF thresholds were calculated as follows:

$$mean(scVAF_{WT\text{-}CTRL}) : 0.000719$$

$$SD(scVAF_{WT\text{-}CTRL}) : 0.000472$$

$$scVAF\ threshold\ for\ calling\ a\ cell\ WT = 0.000719 + 3 \times 0.000472 = 0.002136$$

$$scVAF\ threshold\ for\ calling\ a\ cell\ mutant = 0.002136 + 0.01 = 0.012136$$

We validated this strategy for setting thresholds by bootstrap resampling of the WT control cell scVAFs and fitting each sample to a beta distribution. The resulting probability density distribution quantifies the probability that the scVAF of a WT control cell falls within a particular range of values. The probability of WT control cells having a scVAF measurement above the threshold used to call cells mutant (i.e. mean + 3*SD + 0.01) ranged from 0 to 0.006. Thus, the likelihood of calling false positives (i.e. incorrectly calling a WT cell mutant) is low with this strategy.

Furthermore, we required a minimum number of 10 mutant reads for a cell to be called mutant. Cells with a borderline scVAF (between the WT and mutant scVAF thresholds), where the number of mutant reads was < 10, or where allelic dropout (see next section) was confirmed by analysis of a germline SNP, were called ''undetermined'' and were excluded from downstream analyses comparing genotypes, due to the uncertainty regarding their genotype.

The above strategy was performed separately for gDNA and cDNA amplicons. For each locus, genotyping information from gDNA and cDNA amplicons were then combined, and a consensus genotype assigned. Consensus genotypes were assigned as follows:

1. If the mutation was identified in either the gDNA or cDNA amplicon, the cell was called ''mutant''.
2. If both amplicons were WT, the cell was called ''WT''.
3. If the gDNA amplicon was WT but the cDNA amplicon was undetected or undetermined, the cell was called ''WT''.
4. If the cDNA amplicon was WT but the gDNA amplicon was undetected or undetermined, the cell was called ''undetermined'', due to the high allelic dropout rate of cDNA amplicons.

Allelic dropout (ADO) detection using germline heterozygous SNPs

Cells that appear WT for a CH mutation can either be genuinely WT (where both alleles have been sampled) or there may have been allelic dropout (ADO) of the allele harboring the mutation (where only the WT allele has been sampled). Germline single nucleotide polymorphisms (SNP) were present within the gDNA amplicon for four CH mutations that we analyzed. These SNPs were all heterozygous in the CH samples, but homozygous in the WT-control sample. Since both alleles are present in every cell, these SNPs can be used to determine whether both alleles have been sampled in cells from CH samples, or whether ADO has occurred.

We calculated the scVAF of the SNP with the following formula:

$$SNP\ scVAF = reads\ with\ alternate\ allele\ /\ (reads\ with\ alternate\ allele + reads\ with\ reference\ allele)$$

We defined the allele observed in the WT control sample as the ''reference'' (Ref) allele for each SNP, whereas the allele observed only in the CH sample was the ''alternate'' (Alt) allele. Thus, WT control cells were homozygous Ref (SNP scVAF close to 0), whereas CH sample cells could be heterozygous, homozygous Ref (where there was ADO of the Alt allele) or homozygous Alt (where there was ADO of the Ref allele).

First, we established whether the SNP Alt allele and the CH mutation were found on the same allele (i.e. they were in-phase), or on opposite alleles (i.e. they were out-of-phase), by examining the correlation between SNP and CH mutation scVAFs (Figures S3E and S3F). A positive correlation indicated the SNP Alt allele and the CH mutation were in-phase, while a negative correlation indicated they were out-of-phase.

Next, we called SNP genotypes for each single cell. scVAF thresholds for calling SNP genotypes were determined in a similar way as for CH mutation calling, based on the SNP scVAF distribution in cells from the WT control sample (Figure S3G). Since the control sample was homozygous reference (Hom Ref) for each SNP, this enabled the error rate at each locus to be determined. The SNP scVAF threshold for detection of the SNP Alt allele was calculated as follows:

$$scVAF\ threshold\ for\ detection\ of\ the\ SNP\ Alt\ allele = mean(scVAF_{WT\text{-}CTRL}) + 3 \times SD(scVAF_{WT\text{-}CTRL}) + 0.01$$

Cells with SNP scVAF below this threshold were called Hom Ref for the SNP (i.e. ADO of the SNP Alt allele had occurred).

The SNP scVAF threshold above which a cell was called homozygous alternate (Hom Alt) for the SNP was calculated as the inverse of the lower threshold:

$$SNP\ Hom\ Alt\ scVAF\ threshold = 1 - scVAF\ threshold\ for\ detection\ of\ the\ SNP\ alternate\ allele$$

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Cells with scVAF above this threshold are called Hom Alt (i.e. ADO of the SNP Ref allele had occurred), while cells with VAF between the two thresholds were called heterozygous (Het; i.e. biallelic detection).

We then used the SNP genotype calls and the phasing information to establish whether the allele harboring the CH mutation was detected:

- When a cell was called "mutant" for the CH mutation, the SNP analysis is not required.
- When the SNP Alt allele and the CH mutation were in-phase, cells that appeared WT for the CH mutation, and Hom Ref for the SNP were called "undetermined" (because ADO of the mutant allele had occurred), whereas cells that were Het or Hom Alt for the SNP were called "WT" for the CH mutation (because the allele containing the CH mutation was detected).
- When the SNP Alt allele and the CH mutation were out-of-phase, cells that appeared WT for the CH mutation, and Hom Alt for the SNP were called "undetermined" (because ADO of the mutant allele had occurred), whereas cells that were Het or Hom Ref for the SNP were called "WT" for the CH mutation, mutation (because the allele containing the CH mutation was detected).

Undetermined cells were excluded from downstream analyses between genotypes (given that the allele containing the mutation was not captured due to ADO).

For each allele, we then calculated the fraction of cells in which ADO occurred, by dividing the number of cells in which ADO of one allele occurred by the total number of sampled cells which passed genotyping quality control (Figure S3H).

We note that ADO of the WT allele has no impact on downstream analysis. As all the CH mutations genotyped in this study are heterozygous (apart from the hemizygous *BCORL1* locus), cells in which ADO of the WT allele occurred (which leads to a heterozygous mutant cell being called homozygous mutant) were retained for downstream analyses. We do not distinguish between heterozygous and homozygous mutant calls for CH mutations in downstream analyses comparing genotypes.

ADO estimation using homozygous mutant frequencies

For the loci which did not contain a germline heterozygous SNP in the gDNA amplicon, a direct measure of ADO was not possible. If CH mutations at these loci are assumed to be heterozygous, homozygous mutation calls arise when there is ADO of the WT allele. The WT ADO rate can therefore be estimated by quantifying the fraction of mutant cells that are called homozygous mutant. To identify homozygous mutant cells, we used the inverse of the scVAF threshold used for calling mutant cells (Figure S3I):

$$scVAF\ threshold\ for\ calling\ a\ cell\ homozygous\ mutant = 1 - scVAF\ threshold\ for\ calling\ a\ cell\ mutant$$

Next, we calculated the fraction of cells in which we inferred ADO of the WT allele occurred, with the following formula:

$$WT\ ADO\ rate = n^o\ homozygous\ mutant\ cells \,/\, (n^o\ heterozygous\ mutant\ cells + 2 \times n^o\ homozygous\ mutant\ cells)$$

We first validated this strategy for the 4 loci in which the germline SNP was also present, by comparing WT ADO rates estimated in this way with those calculated with the SNP analysis. Importantly, ADO rates were comparable between the two strategies. Notably, our analysis of germline SNPs also showed that ADO rates for each allele are similar at a given locus (i.e. WT ADO and mutant ADO rates for a single locus are similar; Figure S3H). Hence, the WT ADO rate is a reasonable estimate of the ADO rate for the mutant allele.

For mutations where both gDNA and cDNA genotyping amplicons were used, the WT ADO rate was calculated based on the frequency of cells that were homozygous mutant in both gDNA and cDNA amplicons readouts. Estimated ADO percentages at each locus using this strategy are shown in Figure S3J.

Alternative mutation calling strategy using the minima of the VAF distribution

An alternative strategy for setting thresholds for calling mutant cells based on the scVAF distribution obtained for heterozygous germline SNPs. Plotting the distribution of the scVAFs at a given locus across all genotyped cells results in a histogram containing three peaks: one centered at scVAF ~0.5 and two at values closer to 0 and 1. There will be two minima in the distribution between the three peaks, which can be interpreted as the point that maximally separates heterozygous cells from homozygous cells. In our data, these minima are located at scVAFs equal to ~0.15 and ~0.85. We benchmarked this strategy in our dataset.

Cells with scVAFs between the two minima are likely to be heterozygous, whereas the genotype of cells with scVAFs outside this range is less certain. There are two possible mechanisms by which cells may have VAFs at the extremes of the distribution; either: a) one allele was not amplified (i.e. ADO) and this represents noise in the sequencing data (e.g. PCR error during library preparation or sequencing), leading to a VAF measurement that is not equal to 0 or 1; or b) both alleles were amplified but there was considerable allelic bias/skewing during PCR amplification.

If the first mechanism was the dominant effect, one would expect cells that are truly homozygous to show a similar distribution of scVAFs in the range between 0 and ~0.15. However, the distribution of scVAFs observed in the homozygous reference WT control sample is narrower for all 4 SNPs, suggesting this is not the case. Extrapolating these data to heterozygous CH mutations suggests that most cells with scVAFs in the range between the upper bound of the WT control reference range (~0.01) and ~0.15 are likely to be heterozygous mutant with biallelic detection.

Nonetheless, the genotype of cells with scVAFs in the range between the upper bound of the WT control reference range and ~0.15 is less certain than for cells with scVAFs > 0.15, and we cannot exclude that there was ADO in some cells within this scVAF range. If thresholds for calling heterozygous detection are set based on the minima of the scVAF distribution, the mean ADO rate for the four SNPs is 14.6% per allele, compared to 9.4% using the thresholds defined by the WT control VAF distribution.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

To further determine whether our conclusions hold if the analysis is limited to only the highest-confidence mutant cells, we re-called genotypes based on a scVAF threshold of 0.15. Using the 0.15 threshold to call mutant cells, a total of 183 cells originally called CH^{MUT} were reassigned as "Undetermined" (5.8% of all CH-mutant cells). We repeated the transcriptomic analyses comparing CH^{MUT} and CH^{WT} HSC/MPPs with the new genotyping calls. We observed similar results as in our original analysis, albeit with reduction in enrichment scores and level of significance (Table S6). This suggests our genotyping calling strategy using WT control cells to set thresholds better distinguishes between genotypes and leads to a lower false negative rate (fewer mutant cells being assigned as "Undetermined").

Inference of clonal hierarchies

In samples with multiple mutations, the pattern of mutational co-occurrence was used to determine clonal structures and assign a clonal identity to each cell as previously described.[99,112] In samples where mutations were mutually exclusive, such as in samples NOC131, NOC117, and NOC115, it was clear that these belonged to independent clones. In cases where mutations co-occur in the same cells, a linear or branching clonal structure may be present. We used infSCITE[113] to determine the phylogenetic tree which represented the statistically most likely course of somatic events. As input, we used the matrix containing the mutational status for each locus in each cell and ran infSCITE with default parameters and '-r 200 -L 10000 -fd 0.01 -ad 0.02 -e 0.2 -p 1000'. We confirmed each phylogenetic tree was consistent with the frequency of cells of each genotype and the clonal size determined by bulk BM sequencing VAF.

The occurrence of ADO means that in some cells, a mutation that is present will not be detected. In some of these cases, we were still able to assign a cell to a clone. For example, in sample NOC002 there were 12 cells in which the *TET2* p.R1261C mutation was detected, but the ancestral *TET2* p.Q726X mutation was not detected. In these cases, we inferred that ADO of the ancestral mutation had occurred, and the cell was assigned to the appropriate daughter clone. In all cases, this was a rare occurrence, consistent with our estimates of the ADO rate.

For all downstream analyses, including differential gene expression, the clone assignment rather than the raw genotype was used to categorize WT and mutant cells.

Analysis of FACS index data

Flow cytometry index data were recorded for each single cell during FACS sorting for TARGET-seq+. Fluorescence values were recorded for forward scatter (FSC), back-scatter (BSC; equivalent to side scatter, SSC), Lineage/live/dead, CD34, CD38, CD117, CD45RA, CD10, CD90, CD123, and CD49f (except for the NOC156 control sample). Index data were matched with single-cell identifiers based on the well coordinate and combined with genotyping calls and other metadata into a unified data set. Virtual FACS gating was performed in R based on the sorting strategy. Gates were set based on populations that were negative for each marker. Cells were labelled as positive or negative for each surface marker, and Boolean logic used to assign an immunophenotypic population label. For example, cells that were $Lin^{-}CD34^{+}CD38^{-}CD10^{-}CD45RA^{-}CD90^{+}$ were labelled as immunophenotypic HSC.

Single-cell transcriptome data pre-processing

Mapping and transcript counting

Transcriptome sequencing data were demultiplexed into FASTQ files for each plate with a unique i7-i5 index combination using bcl2fastq. These files contained reads from up to 384 cells with shared plate indexes. A custom python pipeline was used to further demultiplex and map the sequencing reads. First, reads from each plate were demultiplexed using the 14 bp single-cell barcode sequence in Read1 using cutadapt (v3.4). Concurrently, cDNA reads (Read2) were trimmed for polyA tails, Nextera adapters and low-quality reads. This generated individual FASTQ files with single-ended cDNA reads corresponding to each single-cell barcode. Reads were then mapped to the hg38 reference genome and ERCC92 transcripts with STARsolo (v2.7.10a) using the GENCODE v38 reference gene annotation (filtered to include protein coding genes and long non-coding RNAs), and counts for each gene were obtained using default parameters except the following: '–soloType SmartSeq –soloFeatures GeneFull_Ex50pAS'. Sequencing and mapping quality metrics were calculated with FastQC (v0.11.9), Samtools flagstat (v1.12), MultiQC (v1.11) and the outputs of STAR.

Transcript detection and dropout frequency calculation

For the comparison of transcript detection sensitivity between TARGET-seq and TARGET-seq+ (Figures S1G and S1H), data were first downsampled to 5×10^5 reads per cell to remove differences due to unequal sequencing depth. The number of genes detected per cell was calculated as the sum of genes with at least one assigned read.

For calculation of dropout rates (Figure S1I), data downsampled to 5×10^5 reads per cell were used. A random sample of 16 cells per chemistry were compared for JURKAT and 20 cells per chemistry for HSPC. The dropout frequency for a given gene was calculated as the percentage of cells in which the gene was not detected (normalized counts < 1). Genes were divided into three groups to compare the dropout rate in: a) all expressed genes, defined as genes detected in at least 2 cells by any method; b) frequently expressed genes, defined as genes detected in $> 50\%$ of all cells; and c) lowly expressed genes, defined as genes with a mean of 2–10 normalized counts per cell.

Cell-to-cell correlation analysis

Cell-to-cell correlations for JURKAT cells processed with each method (Figure S1J) were calculated using pairwise Pearson correlations in libraries downsampled to 5×10^5 reads per cell.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Single-cell transcriptome analysis

Quality control, normalization, and variable gene identification

Single-cell transcriptome analysis was performed using the SingCellaR package (v1.2.1, https://github.com/supatt-lab/SingCellaR).[65] Metadata including genotyping and FACS index data were matched with single-cell identifiers based on the plate and well coordinates. Cells meeting the following filtering criteria were included in the analysis: reads assigned to genes > 25,000; genes detected > 2,000 and < 15,000; reads assigned to ERCC transcripts < 50%; reads in mitochondrial genes < 15%. Genes expressed in fewer than 10 cells were removed. Reads were normalized by library size using the pool normalization method with prior clustering from the scran package.[114]

Dimensionality reduction, data integration and clustering

Variable genes were identified by fitting a generalized linear model to the relationship between the mean expression and squared coefficient of variation (CV^2) for the ERCC spike-ins, used to estimate technical noise (using the BrenneckeGetVariableGenes function from the M3Drop package).[115,116] Genes for which the CV^2 exceeded technical noise (FDR < 0.05) were considered variable, excluding mitochondrial and ribosomal genes and ERCC transcripts. This identified 17,324 variable genes which were used for principal components analysis (PCA).

Data integration was performed using Harmony[117] to correct for sample effects, using the sample identifier as the batch, and the top 100 principal components (PCs). The top 100 Harmony-adjusted PCs were then used for Uniform Manifold Approximation and Projection (UMAP) analysis and Louvain graph-based clustering implemented in SingCellaR, with k-nearest neighbors (KNN) equal to 15. Effectiveness of the integration was confirmed by: (a) UMAP visualization pre- and post-integration, to confirm representation of all samples across cell types (Figures S2A and S4A–S4H); (b) comparison of cluster identities across samples, to confirm representation of all samples across clusters, while allowing for unequal distributions between samples as is expected from biological variation; (c) concordance between the cluster assignment of cells and their immunophenotype across samples. For example, we confirmed that the majority of cells in the HSC/MPP cluster consisted of immunophenotypic HSCs and MPPs in all samples, and conversely, the majority of immunophenotypic HSCs and MPPs were assigned to the HSC/MPP cluster.

Clusters were manually annotated based on gene set enrichment of published signatures, immunophenotypic surface marker expression, and expression of canonical marker genes. The SingCellaR 'identifyGSEAPrerankedGenes' function was used to prerank genes obtained from differential gene expression analysis comparing each individual cluster with all other clusters, and gene set enrichment analysis (GSEA) was performed using the fgsea package (v1.20.0) against gene sets obtained from 9 studies that have characterized human hematopoiesis.[61–69] Marker genes differentially expressed in each cluster were identified with the SingCellaR 'findMarkerGenes' function, which uses a non-parametric Wilcoxon test on log-transformed, normalized counts, to compare expression levels, and Fisher's exact test to compare the frequency of cells expressing each gene. Louvain clustering identified 28 clusters, which were collapsed into 23 main clusters based on similarity of GSEA results, marker gene expression and immunophenotype.

The HSC/MPP, LMPP, LMPP cycling, and EMPP clusters were further subclustered using the self-assembling manifolds (SAM) algorithm, using default settings with Harmony-adjusted PCs as input and using the sample identifier as the batch.[90] The resulting SAM-weighted PCA was then used as input to generate the UMAP in Figure 6A and for Louvain clustering, which identified 7 clusters. For consistency with earlier analyses, these SAM-derived cluster assignments for LMPP, LMPP cycling, and EMPP were used throughout the paper, while cells assigned to the HSC1-3 and MPP clusters were labelled HSC/MPP in Figures 2, 3, 4, and 5.

Differential abundance analysis

Differential abundance between sample types (CH vs non-CH) and between mutant and WT cells within CH samples was analyzed using MELD,[118] a single-cell compositional analysis method that quantifies the likelihood of a cellular state appearing in each sample or condition, and using the Bayesian model scCODA[76] to determine statistical significance.

For the comparison between sample types (Figures 2F and 2G), only cells sorted as part of the total Lin^-CD34^+ gate were included (excluding the $Lin^-CD34^+CD38^-$ and $CD34^-$ sorting strategies), to avoid bias introduced by enrichment of $CD38^-$ cells. Sample-associated densities were calculated by running MELD using the Harmony-adjusted PCs as input, and with optimal knn and beta values identified using the MELD parameter search. The mean relative density was calculated using the following formula:

$$\textit{Mean relative density} = \textit{mean}(\textit{Density of CH samples}) \,/\, \textit{mean}(\textit{Density of non-CH samples})$$

Thus, a mean relative density ≥ 1 indicates that the probability of observing a given cell is greater in CH samples compared to non-CH samples, whereas a relative density ≤ 1 indicates that the probability is lower in CH samples compared to non-CH samples. scCODA was run sequentially using all cell types as a reference. No credible differences in abundance between CH and non-CH samples were identified with FDR < 0.2 using any cell type as a reference.

To compare mutant and WT cells within CH samples (Figures 3 and 6), sample and genotype-associated densities were calculated for every genotype by running MELD using the Harmony-adjusted PCs as input, with optimal knn and beta values identified using the MELD parameter search. The relative density of single-mutant and WT cells was then calculated (using L1 normalization to enforce cell-wise sum to be 1), and normalized to the mean relative density in the HSC/MPP cluster for each sample, in order to quantify the relative expansion or contraction of the clone downstream of the HSC/MPP, using the following formula:

$$\textit{Normalized likelihood} = \textit{Relative likelihood of MUT} : \textit{WT} \,/\, \textit{mean}(\textit{Relative likelihood of MUT} : \textit{WT in HSC} \,/\, \textit{MPP})$$

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

These values are visualized per sample in Figures 3D, 3G, 3J, 3N, S4A–S4H, and S6Q. Finally, the normalized likelihood was averaged across samples for visualization in Figures 3C, 3F, and 6G. For the analyses in Figures 3 and S4, scCODA was run using HSC/MPP as the reference cell type. For the analyses in Figure 6, scCODA was run using HSC1 as the reference cell type. Changes classified by scCODA as credible after correction for multiple comparisons (FDR < 0.1 and/or FDR < 0.2) were considered statistically significant.

Pseudotime analysis

Diffusion map embeddings[119] were calculated in scanpy using the Harmony-adjusted PCs as input to the neighborhood graph, excluding the T cell, plasma cell, and endothelial cell clusters. Diffusion pseudotime was then calculated using the HSC at the extreme of the second diffusion component as the root cell. Pseudotime scores were extracted for cells in the HSC/MPP, LMPP, GMP, pDC, and Monocyte clusters and plotted on the UMAP embedding to visualize the myeloid trajectory. For comparison of *TET2*^{MUT} and *TET2*^{WT} cell density along the myeloid trajectory, cells were downsampled to an equal number per sample (n = 176 cells from each of the 4 samples).

Differential gene expression analysis

Differential expression testing was performed with a linear mixed model to account for sample covariance using the dream pipeline from the variancePartition package,[120] which is based on limma-voom.[121] Testing was performed on uncorrected, raw counts, using the scran normalization size factors. Genes were filtered to exclude genes expressed in fewer than 10% of cells in either group, except for pDCs, monocytes, and Pro-B/Pre-B cells, where a 20% filter was used. A linear mixed model was fitted to each gene using 'dream' and differential expression testing was performed using 'variancePartition::eBayes'. For comparisons between sample types, the sample type was used as the test variable, and the sample identifier, age, sex, and batch effects included as covariates. For comparisons between genotypes within CH samples, the clone was used as the test variable, and the sample identifier and batch effect included as mixed effect covariates. Samples were excluded from the comparison if they had less than 5 cells in either genotype, except for pDCs and monocytes, where a minimum of 2 cells was used. P values were adjusted for multiple testing with Benjamini-Hochberg correction, and differentially expressed genes were defined as those with FDR < 0.1. Analyses were corroborated using a pseudo-bulk approach with EdgeR (results not shown). For defining the CH^{WT} HSC/MPP and non-CH HSC/MPP signatures, thresholds of FDR < 0.1 and $\log_2$FC > 0.5 were used. Differential expression results are listed in Table S6.

For additional validation, differential expression analysis was performed between CH^{MUT} and CH^{WT} cells within individual CH samples using WilcoxAUC. Testing was performed on scran normalized counts, which were filtered to exclude genes expressed in fewer than 20% of cells in either group.

To assess the impact of ADO rates on differential expression analysis between CH^{MUT} and CH^{WT} cells, a random 10% sample of CH^{MUT} HSC/MPPs was deliberately re-classified as CH^{WT}, and the differential expression analysis repeated. GSEA was performed on the resulting ranked gene lists and the results summarized in Table S6.

Changes in gene expression along pseudotime

For plotting gene expression along pseudotime, $\log_2$ normalized expression data were fitted to the pseudotime rank using a generalized additive model (GAM) separately for WT and mutant cells.

Gene set enrichment analysis (GSEA)

Gene rankings for gene set enrichment analysis (GSEA) across multiple samples were generated by differential gene expression testing using the dream mixed model, as described above. Genes were ranked by the z statistic from dream. For analyses between CH^{MUT} and CH^{WT} cells within individual CH samples, the gene rankings from wilcoxAUC were used, ranking on the AUC value. To perform GSEA, the fgseaMultilevel function from the fgsea package[122] was used. Gene sets were obtained from MsigDB v7.5.1 and published studies. Hematopoietic gene sets used in GSEA and AUCell analyses relating to Figures 4, 5, and 6 are listed in Table S5. Significantly enriched gene sets were filtered using the FDR as described in each figure.

AUCell signature analysis

The AUCell package (v1.18.1)[82] was used to quantify the gene set activity in single cells. AUCell gene-expression rankings were created using the SingCellaR 'Build_AUCell_Rankings' function. AUCell gene signature enrichment was then calculated using the 'Run_AUCell' function with the gene matrix transposed (GMT) file of gene sets. Hematopoietic gene sets were the same as those used for GSEA analysis described above. Differences in mean AUCell scores between WT and mutant cells were tested by a linear mixed model, using clone identity as the fixed effect and sample identity as mixed effects. P-values were obtained by a likelihood ratio test of the full model with the clone effect against the model without the clone effect.

SCENIC transcription factor regulon analysis

To infer transcription factor (TF) regulon activity, regulon analysis was performed using pySCENIC.[82] pySCENIC was run as per the workflow guidelines from Van de Sande et al.[101] to identify candidate TF-regulons, using the filtered, pre-processed raw counts as the input, and a list of human TFs from Lambert et al.[123] Candidate regulons were pruned using the annotations of TF motifs 'motifs-v10nr_clust-nr.hgnc-m0.001-o0.0.tbl', and CisTarget was applied using the 'mc_v10_clust' databases of known human TF motifs annotated at: a) 500 bp upstream and 100 bp downstream of the transcription start site (TSS); and b) 10 kilobases centered around the TSS. No drop-out masking was applied. Enrichment of refined TF regulons was quantified using AUCell, with default parameters. Tests for differential regulon activity were performed using a linear mixed model, as described above. Additionally, WilcoxAUC was used to test differential regulon activity between mutant and WT cells within individual samples. The mean area under the curve (AUC) was calculated to quantify the mean change in activity across samples.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

Analysis of TCL1A expression

A potential way to assess genotyping accuracy would be to quantify expression of a gene specific for a certain genotype. It has been proposed that *TCL1A* is aberrantly expressed in *TET2*^{MUT} HSCs, and that this effect is reduced by the presence of the rs2887399 variant T allele.[77] We therefore analyzed the frequency of immunophenotypic HSC/MPPs expressing *TCL1A* transcripts across CH and non-CH samples. *TCL1A*^{hi} cells were defined as cells with > 2 reads mapping to *TCL1A*.

If *TCL1A* expression was a specific marker for *TET2*^{MUT} HSC/MPPs, it would be expected that *TCL1A*^{hi} HSC/MPPs are transcriptionally more similar to *TET2*^{MUT} HSC/MPPs than to other *TET2*^{WT} HSC/MPPs from the same sample. We therefore generated *TET2*^{MUT} and *TET2*^{WT} gene signatures by differential expression analysis between *TET2*^{MUT} and *TET2*^{WT} HSC/MPPs from *TET2*^{MUT} CH samples and validated these by AUCell scoring and ROC AUC analysis with WilcoxAUC (Figures S3M–S3O). We then compared the expression of these signatures in *TCL1A*^{hi} *TET2*^{WT} cells against other *TET2*^{WT} cells and *TET2*^{MUT} cells.

FACS sorting and snRNA-seq for 'in-house' aging dataset

BM samples were thawed via slow dropwise addition of X-VIVO 10 media (LONZA) with 50% FBS and 100 μg/mL DNaseI (Roche). Cells were centrifuged at 400 × g for 10 min, then dead cell depleted using a commercial kit (EasySep Dead Cell Removal (Annexin V) Kit, STEMCELL) per the manufacturer's instructions. Cells were resuspended in PBS + 5% FBS and stained for 15 min at RT for fluorescence-activated cell sorting with the following antibodies: anti-CD45RA-FITC (1:50, BD, clone HI100), anti-CD90-PE (1:50, BD, clone 5E10), anti-CD19-BV711 (1:50, BD, clone SJ25C1), anti-CD49f-PE-Cy5 (1:50, BD, clone GoH3), anti-CD271-APC (1:100, Miltenyi, ME20.4-1.H4), anti-CD34-APC-Cy7 (1:200, BD, clone 581), anti-CD38-PE-Cy7 (1:200, BD, clone HB7), anti-CD10-AlexaFluor700 (1:50, BD, clone HI10a), anti-CD14-BV605 (1:200, BD, clone M5E2), anti-CD45-V500 (1:50, BD, clone HI30) and anti-CD33-BV421 (1:100, BioLegend, clone WM53). Cells were washed following staining and resuspended in PBS + 2% FBS containing propidium iodide and filtered through 40μm nylon mesh for cell sorting. Lin$^-$CD34$^+$CD38$^-$ and Lin$^-$CD34$^+$CD38$^+$ populations were sorted into PBS + 0.04% BSA + EDTA on a BD FACSAria Fusion or BD FACSAria III. Cells were counted and Lin$^-$CD34$^+$CD38$^-$ and Lin$^-$CD34$^+$CD38$^+$ cells mixed in the following manner (1:0.33 for 24yM, 1:1 for 26yF, 1:1 for 70yF, and 1:0.5 for 77yF) for downstream 10x Genomics multiome sample preparation by the Princess Margaret Genome Centre.

Single-nucleus RNA-seq processing - In-house aging dataset

Single-nucleus RNA-seq processing was performed using Seurat 4.3.0 in R and applied to each sample before merging. The UMI count matrix (BM24M, BM26F, BM70F, and BM77F) was loaded in the R environment using Read10X. Doublets were identified using scDblFinder on the RNA with basic parameters after filtering genes expressed in more than 3 cells, cells with more than 200 features and more than 0.05 percent ribosomal genes. After removing doublets, cells that passed the following filtering criteria were used for downstream analysis: 200 < unique feature counts < 3000, percent mitochondrial genes < 10%. The 10x count matrix for each sample was corrected for ambient RNA contamination using SoupX and used for downstream analysis with the cells that passed quality control. The samples were merged, and the counts were normalized using Scran. The 2000 highly variable features were selected using the "vst" selection method with FindVariableFeatures in Seurat. The cells were scaled, and the samples were integrated using Harmony correcting the sample assignments as a covariate. The optimal number of Harmony-corrected PCA components for downstream analysis was assessed using an elbow plot optimizing at 10. A k-nearest neighbors graph was constructed using FindNeighbors with the Harmony corrected principal components (PCA), and clusters were identified using the Louvain algorithm (resolution = 0.8). A UMAP was constructed using the RunUMAP function at 30 neighbours and 10 Harmony corrected PCA components.

Single-cell RNA-seq processing - Ainciburu et al. dataset

The single-cell RNA-seq data was downloaded from GSE18029886, processed using Seurat 4.3.0 in R, and applied to each sample before merging. The UMI count matrix (young1, young2, young3, young4, young5, elderly1, elderly2, elderly3) was loaded in the R environment using Read10X. Doublets were identified using scDblFinder on the RNA after filtering genes expressed in more than 3 cells, cells with more than 200 features and more than 0.05 percent ribosomal genes. After removing doublets, cells passing unique feature count (nFeature_RNA) and percent mitochondrial gene thresholds (percent.mt) were used for downstream analysis (young1: 200 < nFeature_RNA < 4000, percent.mt < 10; young2: 200 < nFeature_RNA < 2700, percent.mt < 10; young3: 200 < nFeature_RNA < 4000, percent.mt < 5; young4: 200 < nFeature_RNA < 4000, percent.mt < 5; young5: 200 < nFeature_RNA < 5000, percent.mt < 10; elderly1: 200 < nFeature_RNA < 4000, percent.mt < 10; elderly2: 200 < nFeature_RNA < 4000, percent.mt < 10; elderly3: 200 < nFeature_RNA < 5000, percent.mt < 10). The samples were merged and normalized using Scran. The 2000 highly variable features were selected using the "vst" selection method with FindVariableFeatures in Seurat. The cells were scaled, and the samples were integrated using Harmony correcting the sample assignments and technology (10x 3' V2 chemistry, 10x 3' V3 chemistry) as covariates. The optimal number of Harmony-corrected PCA components for downstream analysis was assessed using an elbow plot optimizing at 15. A k-nearest neighbors graph was constructed using FindNeighbors with the Harmony reduction, and clusters were identified using the Louvain algorithm (resolution = 0.5). A UMAP was generated using the RunUMAP function at 30 neighbours and 15 PCA components.

Please cite this article in press as: Jakobsen et al., Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging, Cell Stem Cell (2024), https://doi.org/10.1016/j.stem.2024.05.010

CellPress
OPEN ACCESS

Single-cell RNA-seq processing - Aksöz et al. dataset

This dataset consists of 10x 3' V2 single-cell RNA-seq data from FACS-purified $Lin^{-}CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}$ HSCs from 3 young and 3 aged donors (all male).[85] Briefly, the raw FASTQ files were aligned against the GRCh38 (Ensembl 93) reference genome (10X Cell Ranger reference GRCh38 v3.1.0) and quantified using the Cell Ranger pipeline (v3.1.0) with default parameters and further processed using Seurat (v4.3.0). Quality control was performed separately for each donor by first filtering out cells with < 200 genes detected, and then retaining only cells with < 10% mitochondrial reads and gene counts that are less than double the median gene count detected in the data for that donor. Genes detected in less than 3 cells were removed. All cells that passed quality control were included in differential expression analysis.

Aged vs Young HSC Differential Expression

Pseudobulk profiles of HSCs from each donor were created by taking the sum of all counts for each gene across cells belonging to the HSC cluster within that donor. For the in-house aging dataset, raw counts from young and aged HSC pseudobulks were modeled with DESeq and differential expression was run between aged HSC and young HSC with donor sex as a covariate. Young HSC and aged HSC-specific genes with $\log_2$FoldChange > 1 and FDR < 0.01 were retained as signatures for downstream analysis. For the Ainciburu dataset, DESeq was run on raw counts from young and aged HSC pseudobulks only within samples profiled by 10x 3' scRNA-seq V2 chemistry to avoid technology-driven batch effects. This comparison in the Ainciburu et al. dataset was confounded by donor sex, wherein all aged samples were male, and all young samples were female. To attenuate this, sex specific genes (X-inactivation genes XIST and TSIX, as well as ChrY genes outside of the para-autologous region) were filtered out from the DE results. Young HSC and aged HSC-specific genes with $\log_2$FoldChange > 1 and FDR < 0.01 were retained as signatures for downstream analysis. For the Aksöz et al. dataset, raw count pseudobulks were modeled with EdgeR as implemented in the Libra (v1.0.0) package,[124] and differential expression was run using a likelihood ratio test between aged HSC and young HSC. Gene identifiers were converted to GENCODE v38 and young HSC and aged HSC-specific genes with $\log_2$FoldChange > 1 and FDR < 0.01 were retained as signatures for downstream analysis, excluding genes not in the GENCODE reference annotation. Differential expression results are listed in Table S6.

The quality of each resulting signature was evaluated by scoring across donors within our CH cohort and evaluating their association with age. While we validated that aged HSC signatures from each dataset were positively correlated with age, young HSC signatures were uncorrelated with age, rather than having the expected negative correlation. Thus, only aged HSC signatures were used for downstream analysis (Table S5).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis and statistical tests were performed using R version 4.2.1. Plots were generated using ggplot2 (v3.3.6) or FlowJo (v10.8.1). Detail on statistical tests used in the different figures and definition of relevant summary statistics are included in each figure legend.

Cell Stem Cell, Volume *31*

Supplemental Information

Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging

Niels Asger Jakobsen, Sven Turkalj, Andy G.X. Zeng, Bilyana Stoilova, Marlen Metzner, Susann Rahmig, Murtaza S. Nagree, Sayyam Shah, Rachel Moore, Batchimeg Usukhbayar, Mirian Angulo Salazar, Grigore-Aristide Gafencu, Alison Kennedy, Simon Newman, Benjamin J.L. Kendrick, Adrian H. Taylor, Rasheed Afinowi-Luitz, Roger Gundle, Bridget Watkins, Kim Wheway, Debra Beazley, Alex Murison, Alicia G. Aguilar-Navarro, Eugenia Flores-Figueroa, Stephanie G. Dakin, Andrew J. Carr, Claus Nerlov, John E. Dick, Stephanie Z. Xie, and Paresh Vyas

SUPPLEMENTAL FIGURES AND LEGENDS

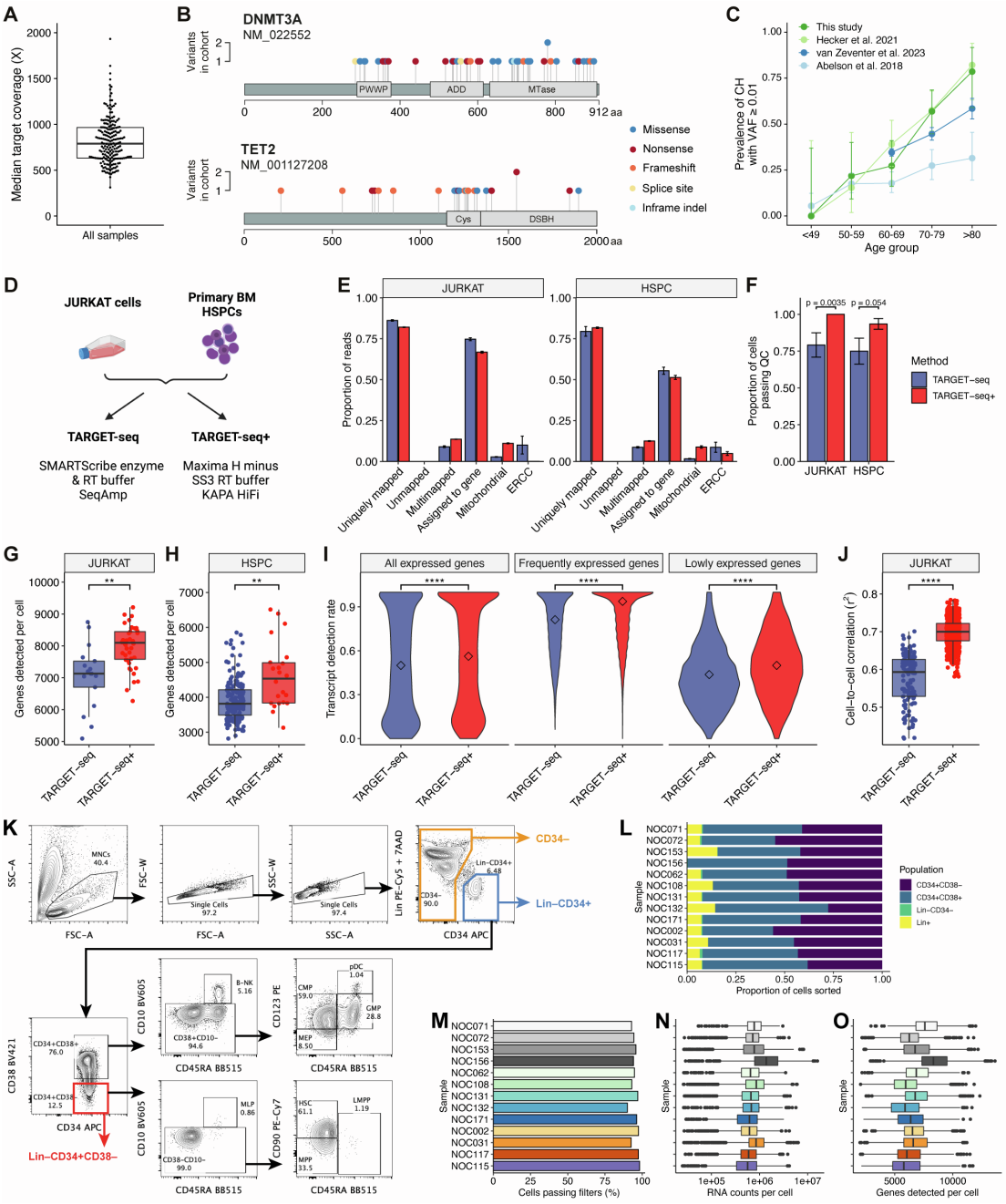


Figure S1. Identification of cases with age-related CH in individuals undergoing hip replacement surgery, validation of TARGET-seq+, and application of TARGET-seq+ to CH and WT samples, related to Figures 1 and 2 and STAR Methods.

(A) Median coverage across the target region (97 genes in total) in targeted DNA sequencing of BM samples. Each point represents a sample. The boxplot shows the median and interquartile range.

(B) Distribution of mutations in coding regions of *DNMT3A* and *TET2*. Positions of protein domains are shown: PWWP, proline-tryptophan-tryptophan-proline; ADD, ATRX-DNMT3-DNMT3L; MTase, cytosine methyltransferase; Cys, cysteine-rich; DSBH, double-stranded β-helix. Mutations are colored by the predicted effect on the protein.

(C) Prevalence of CH with at least one driver mutation (VAF ≥ 0.01) by age. BM DNA sequencing data from participants in this study (n = 195) are compared with another hip replacement cohort (Hecker et al.[1]; n = 109 BM and n = 91 PB samples), and with two other studies of the general population (Abelson et al.[2]; n = 676 PB samples, and van Zeventer et al.[3]; n = 3,359 PB samples). Error bars represent 95% confidence intervals. Studies performed on individuals undergoing hip replacement surgery are indicated in green, while other studies are indicated in blue colors.

(D) Experimental scheme for validation of TARGET-seq+ in JURKAT cells and primary human $CD34^+$ HSPCs. Metrics were compared to single cell transcriptome libraries generated with the original TARGET-seq protocol.

(E) Sequencing statistics of single cell transcriptome libraries from JURKAT and primary $CD34^+$ HSPCs processed with TARGET-seq and TARGET-seq+. Data are represented as the proportion ± SE.

(F) Proportion of cells passing quality control (QC) for transcriptome libraries, for each method (n = 24 JURKAT for TARGET-seq; 46 JURKAT for TARGET-seq+; 24 HSPC for TARGET-seq; and 46 HSPC for TARGET-seq+). Data are represented as the proportion ± SE. P-values calculated by Fisher's exact test.

(G and H) Number of genes detected per cell in JURKAT cells (G) and primary $CD34^+$ HSPCs (H). In both cases, reads were downsampled to 5×10^5 reads per cell. Each dot represents a cell, and each boxplot represents the median and first and third quartiles. P-values calculated by unpaired two-tailed *t*–test. **p < 0.01.

(I) Comparison of transcript detection rates between TARGET-seq and TARGET-seq+ for JURKAT cells. A subsample of 16 cells per chemistry was analyzed. Reads were downsampled to 5×10^5 reads per cell. Dropout frequencies are shown for all genes expressed in at least 2 cells (left), for genes expressed in > 50% of cells by any method (middle), and for genes with low expression (mean normalized counts of 2-10 per cell; right). Diamonds show the median detection rate for each condition. P-values calculated by two-tailed Wilcoxon rank-sum test.

(J) Quantification of the reproducibility of gene expression in JURKAT cells, for TARGET-seq (n = 17 cells) and TARGET-seq+ (n = 43 cells). r^2 values for all pairwise cell-to-cell Pearson's correlations in libraries downsampled to 5×10^5 reads per cell are shown. P-values calculated by unpaired two-tailed *t*–test. ****p < 0.0001.

(K) FACS index sorting strategy for TARGET-seq+ experiments. Sorting gates for $Lin^{-}CD34^{+}$ (blue), $Lin^{-}CD34^{+}CD38^{-}$ (red), and $CD34^{-}$ (orange) cells are shown.
(L) Proportion of cells obtained in each immunophenotype for TARGET-seq+ analysis in each sample.
Panels (M) to (O) show TARGET-seq+ transcriptome quality control and sequencing statistics by sample. In panels (N) and (O), boxplots show the median and interquartile range.
(M) Percentage of cells passing quality control filters.
(N) Sequencing depth (RNA counts) per cell.
(O) Number of genes detected per cell.

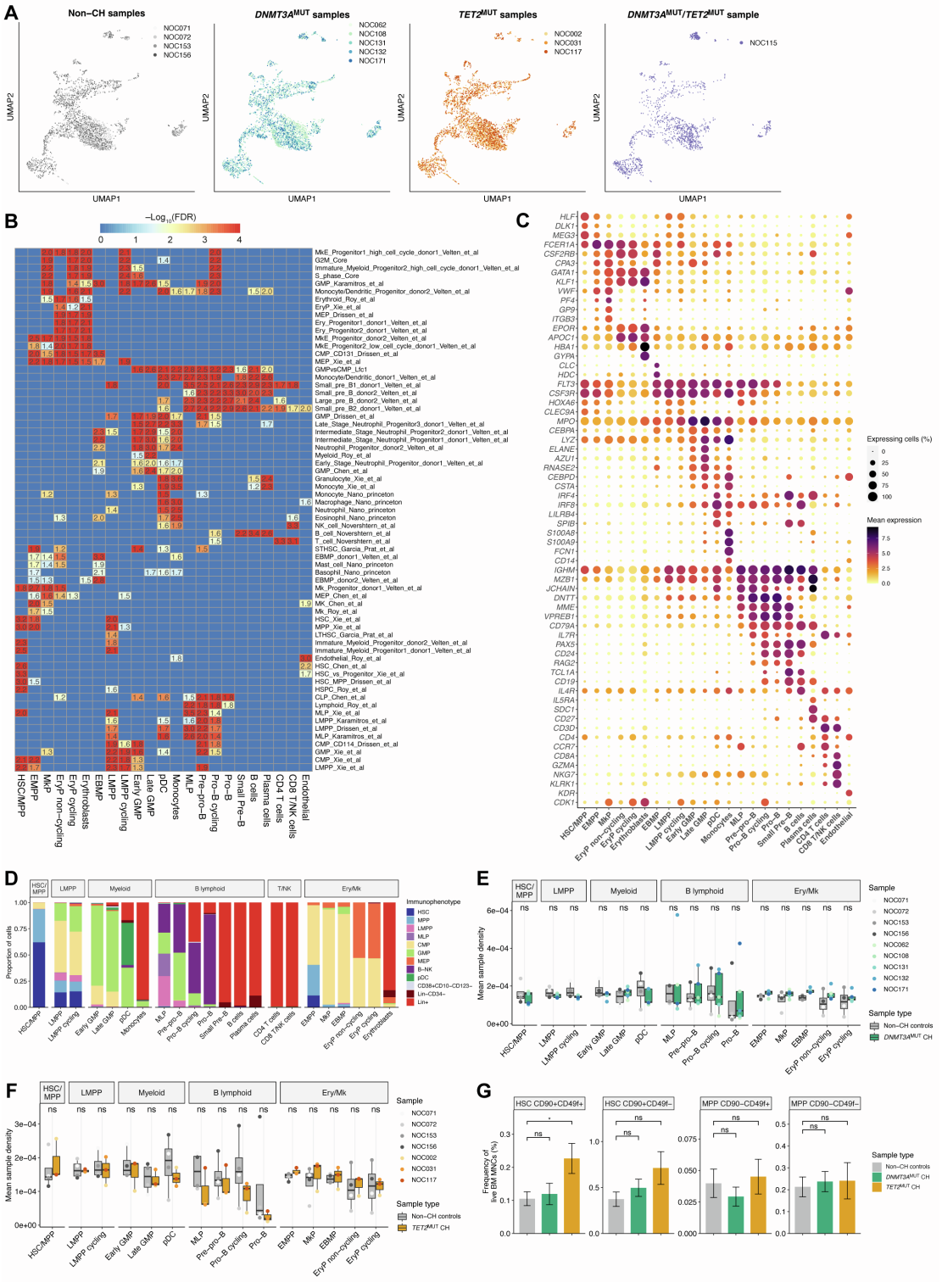


Figure S2. TARGET-seq+ applied to *DNMT3A*- and *TET2*-mutant clonal hematopoiesis, related to Figure 2 and STAR Methods.

(A) UMAPs colored by sample, split by sample genotype. Colors indicate cells from different samples.

(B) Heatmap of GSEA results for genes differentially expressed between each cluster and all other clusters, using published gene signatures for HSPCs.[4-12] Colors represent $-\log_{10}$ of the false discovery rate (FDR). Numbers inside the heatmap indicate the normalized enrichment score for each comparison, showing only those with FDR < 0.05.

(C) Expression of marker genes by cluster. The size of each dot represents the percentage of cells expressing each gene and the color represents the mean expression level (log normalized counts).

(D) Proportions of each immunophenotypic population in each transcriptional cluster (x-axis).

(E) Boxplots summarizing Figure 2F, showing the mean density of cells per cluster in *DNMT3A*^{MUT} CH samples (green) compared to WT control samples (grey). Median and interquartile range are shown. Each dot represents the mean value from a single sample (color legend on the right). No credible differences between *DNMT3A*^{MUT} CH samples and control samples were identified (with FDR < 0.2) using scCODA.

(F) As in (E) summarizing Figure 2G, comparing *TET2*^{MUT} CH samples (orange) with WT control samples (grey). No credible differences between *TET2*^{MUT} CH samples and control samples were identified (with FDR < 0.2) using scCODA.

(G) Barplots showing the frequency of CD90$^{+/-}$ and CD49f$^{+/-}$ HSC/MPP cells as a percentage of total BM MNCs. Data are represented as mean ± SEM. P-values calculated by Wilcoxon rank-sum test with Holm-Bonferroni multiple testing correction. *p < 0.05.

HSC, hematopoietic stem cell; MPP, multipotent progenitor; EMPP, erythroid/megakaryocyte-primed multipotent progenitor; MkP, megakaryocytic progenitor; EryP, erythroid progenitor; EBMP, Eosinophil-basophil-mast cell progenitor; LMPP, lymphoid-primed multipotent progenitor; GMP, granulocyte-monocyte progenitor; pDC, plasmacytoid dendritic cell progenitor; MLP, multi-lymphoid progenitor; B-NK, B and NK cell progenitor.

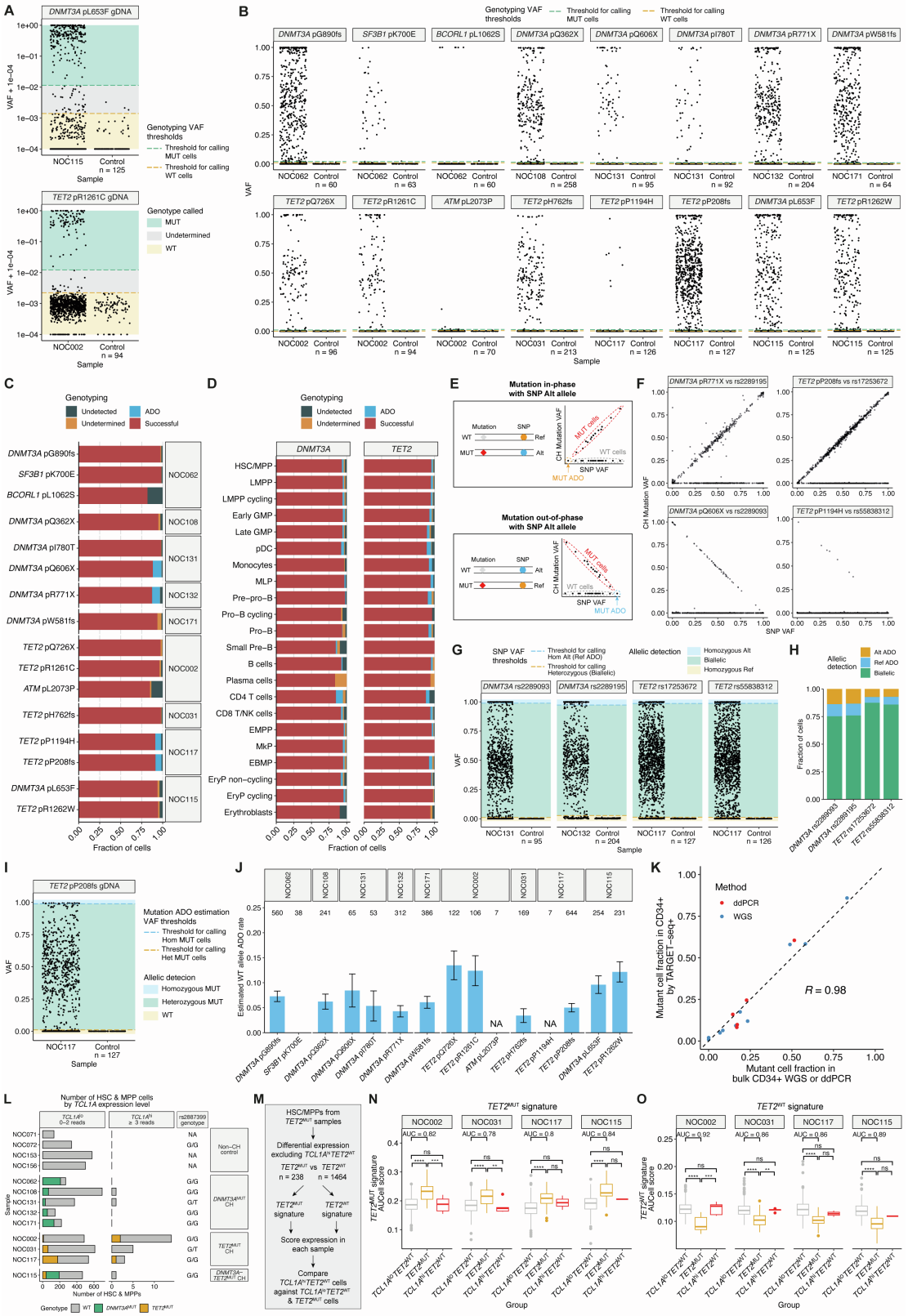


Figure S3. TARGET-seq+ single-cell genotyping quality control and validation, related to Figure 3 and STAR Methods.

(A) Single-cell variant allele frequency (scVAF) distribution (y-axis; logarithmic scale) obtained from gDNA amplicons for two representative mutations in *DNMT3A* and *TET2*, comparing CH samples (on the left of each graph) versus the WT control sample (on the right of each graph). Each dot represents a single cell. Cells below the lower threshold (orange) are called WT; cells between the two thresholds (grey) are called undetermined; cells above the upper threshold (green) are called mutant (detailed in STAR Methods).

(B) Plots of scVAF distribution (y-axis, linear scale) from gDNA amplicons across all mutant loci, comparing CH samples (on the left of each graph) versus WT control samples (on the right of each graph). Each dot represents a single cell. Thresholds for calling cells mutant or WT are shown.

(C) Fraction of cells successfully genotyped (red) for each mutation and sample analyzed. In 4 amplicons, allelic dropout (ADO) of the mutant allele (blue) could be identified by analysis of a germline heterozygous SNP in the same amplicon (STAR Methods). Cells with borderline VAF were assigned an undetermined genotype (orange) for that mutation.

(D) Fraction of cells successfully genotyped for *DNMT3A* and *TET2* across transcriptionally defined cell types.

(E) Schematic overview of how germline heterozygous SNPs may be phased with mutations in the same genotyping amplicon. When the SNP Alternate (Alt) allele is in-phase with the mutation (i.e. they are located on the same allele, top), cells where there has been ADO of the mutant allele will have a single-cell variant allele frequency (scVAF) close to 0% for both the mutation and the SNP. When the SNP Alt allele is out-of-phase with the mutation (i.e. they are located on opposite alleles, bottom), cells where there has been ADO of the mutant allele will have a scVAF of ~0% for the mutation but a scVAF of ~100% for the SNP.

(F) scVAF of mutations (y-axis) plotted against the scVAF of germline heterozygous SNPs (x-axis) found in the same amplicon. Each dot represents a single cell. The SNP and the mutation are either in-phase, i.e. they are located on the same allele (top plots), or they are out-of-phase, i.e. they are located on opposite alleles (bottom plots), as evidenced by the correlation in scVAF in mutant cells.

(G) Plots of scVAF distribution (y-axis; linear scale) across 4 germline heterozygous SNPs used for ADO detection in gDNA amplicons (STAR Methods). Each dot represents a single cell. Cells from the CH samples are plotted on the left and cells from the WT control sample are plotted on the right of each graph. The WT control sample was homozygous reference for all 4 SNPs. Cells below the lower threshold (orange) are called Homozygous Ref; cells between the two thresholds (green) are called heterozygous (Biallelic detection); cells above the upper threshold (blue) are called Homozygous Alt.

(H) Frequency of biallelic and mono-allelic detection of 4 germline heterozygous SNPs co-amplified with somatic mutations.

(I) Plot of scVAF distribution as in (B), but for a representative *TET2* mutation, showing how homozygous mutant cells were called. Cells above the lower threshold (orange) are called mutant (MUT); cells above the upper threshold (blue) are called Homozygous MUT. The threshold for calling cells homozygous mutant is the inverse of the lower threshold. The frequency of cells called Homozygous MUT was used to estimate ADO rates (see panel J).
(J) Bar plots showing the estimated WT ADO rate for each mutant locus. The hemizygous *BCORL1* locus was excluded. Samples are indicated on the top. Numbers above each bar show the number of mutant cells analyzed for each locus. Error bars indicate the standard error. NA indicates those mutations for which the ADO rate was not calculated due to the small number of mutant cells.
(K) Comparison of the mutant cell fraction for each mutation identified in $Lin^{-}CD34^{+}$ HSPCs by TARGET-seq+ (y-axis) with the clonal cell fraction in bulk $Lin^{-}CD34^{+}$ HSPCs (x-axis) as measured by whole genome sequencing or ddPCR. *R* indicates the Pearson correlation coefficient.
(L) Number of immunophenotypic HSC & MPP cells in non-CH and CH samples according to *TCL1A* expression level. Bars are colored by cell genotype. Where available, the rs2887399 genotype is shown for each sample. NA, not available.
(M) Strategy for analysis of genotype-specific expression profile in $TCL1A^{hi}TET2^{WT}$ HSC/MPP. $TET2^{MUT}$ and $TET2^{WT}$ signatures were defined based on differential expression analysis between $TET2^{MUT}$ and $TET2^{WT}$ HSC/MPPs from the $TET2^{MUT}$ CH samples (excluding $TCL1A^{hi}TET2^{WT}$ cells).
(N-O) AUCell scores for the $TET2^{MUT}$ (N) and $TET2^{WT}$ (O) signatures in immunophenotypic HSC/MPPs from $TET2^{MUT}$ CH samples. Two populations of $TET2^{WT}$ cells are shown: $TCL1A^{hi}TET2^{WT}$ cells (defined as WT cells with > 2 *TCL1A* reads) and $TCL1A^{lo}TET2^{WT}$ (defined as WT cells with ≤ 2 *TCL1A* reads). Performance of each signature in predicting $TET2^{MUT}$ vs $TET2^{WT}$ genotype is shown by the area under the ROC curve (AUC). P-values calculated by two-tailed Wilcoxon rank-sum test. **p < 0.01, ***p < 0.001, ****p < 0.0001.

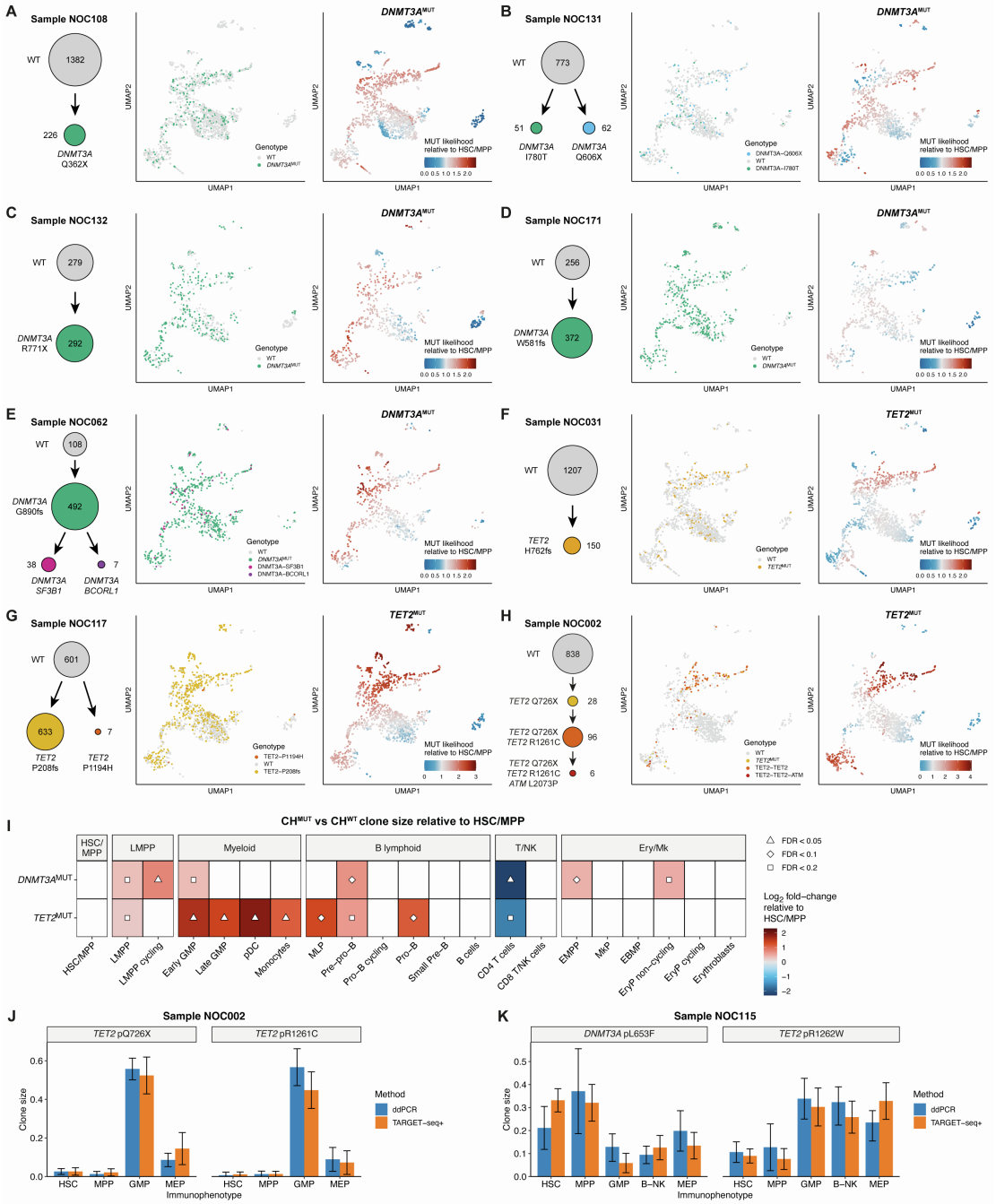


Figure S4. Sample-specific patterns of clonal expansion of *DNMT3A*- and *TET2*-mutant clones, related to Figure 3.

(A) to (H) Clonal structures and patterns of clonal expansion for individual *DNMT3A*- and *TET2*-mutant CH samples. For each sample, the clonal structure determined by single-cell genotyping is shown on the left. Numbers of cells assigned to each genotype are shown. The middle panels show successfully genotyped cells from each sample on the UMAP embedding, labelled by cell genotype. Right-hand panels show the relative likelihood for mutant clones

carrying a single mutation in *DNMT3A* or *TET2* on the UMAP. In each case, the mutant clone likelihood is normalized to the mean likelihood in the HSC/MPP cluster. Values > 1 indicate clonal expansion relative to HSC/MPP and values < 1 indicate smaller clone size relative to HSC/MPP.

(I) Heatmap of $\log_2$FC in abundance of mutant clones relative to the HSC/MPP as estimated with scCODA. Only significant results at an FDR < 0.2 are shown, with all nonsignificant differences plotted as white. Symbols (legend on the right) indicate whether results are significant at FDR threshold of 0.05, 0.1 or 0.2.

(J) Mutant clone size for two *TET2* mutations in sample NOC002 within immunophenotypic populations. Data from ddPCR of FACS-sorted populations (blue bars) is compared with frequencies of mutant cells in TARGET-seq+ data (orange bars). Error bars represent 95% confidence intervals.

(K) Same as in (J) but for the *DNMT3A* and *TET2* mutations in sample NOC115.

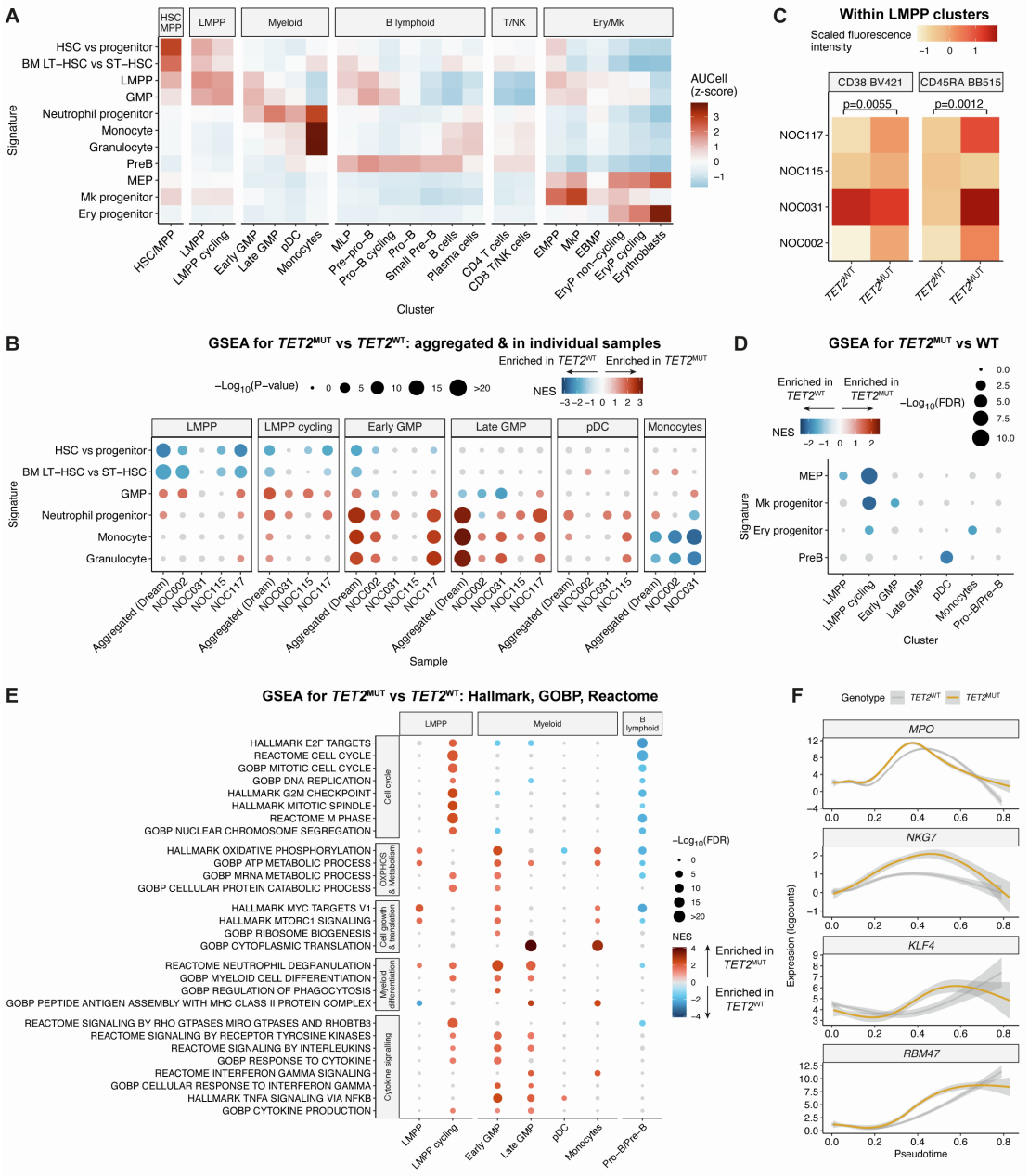


Figure S5. Consequences of *TET2* mutations in myeloid progenitors, related to Figure 4.

(A) Heatmap of scaled AUCell expression scores for HSC, myeloid, B-lymphoid, megakaryocytic, and erythroid gene signatures (y-axis) used in Figures 4C, 4D, S5B and S5D, by cluster (x-axis). See also Table S5.

(B) Comparison of GSEA results between the aggregated Dream linear mixed model and differential expression analysis in individual samples using a Wilcox test. GSEA was performed against hematopoietic signatures comparing *TET2*^{MUT} versus *TET2*^{WT} cells within each LMPP and GMP cluster. Results are not shown for samples with < 3 cells in either

genotype in a particular cluster. Color intensity indicates the normalized enrichment score (NES); positive NES values indicate enrichment in mutant cells. Signatures with P-value > 0.05 are colored grey.

(C) Expression of CD38 and CD45RA cell surface protein, measured by FACS indexing in *TET2*^{WT} and *TET2*^{MUT} cells, in the LMPP clusters. P-values calculated by linear mixed model.

(D) GSEA against erythroid, megakaryocytic, and lymphoid gene signatures, comparing *TET2*^{MUT} against *TET2*^{WT} cells within each lympho-myeloid cluster. Pre-pro-B, Pro-B, and Small Pre-B cells were combined for the Pro-B/Pre-B analysis. Differential expression analysis was performed accounting for sample and batch effects. Cells from the 4 *TET2*^{MUT} CH samples were included. Signatures with FDR > 0.05 are colored grey. Positive NES values indicate enrichment in mutant cells.

(E) GSEA against Hallmark, Gene Ontology biological process (GOBP), and Reactome signatures comparing *TET2*^{MUT} against *TET2*^{WT} cells within the LMPP, GMP, pDC, monocyte, and Pro-B/Pre-B clusters. Signatures with FDR > 0.05 are colored grey. Positive NES values indicate enrichment in mutant cells.

(F) Local regression of gene expression values along myeloid pseudotime for myeloid lineage-affiliated genes, comparing *TET2*^{MUT} and *TET2*^{WT} cells.

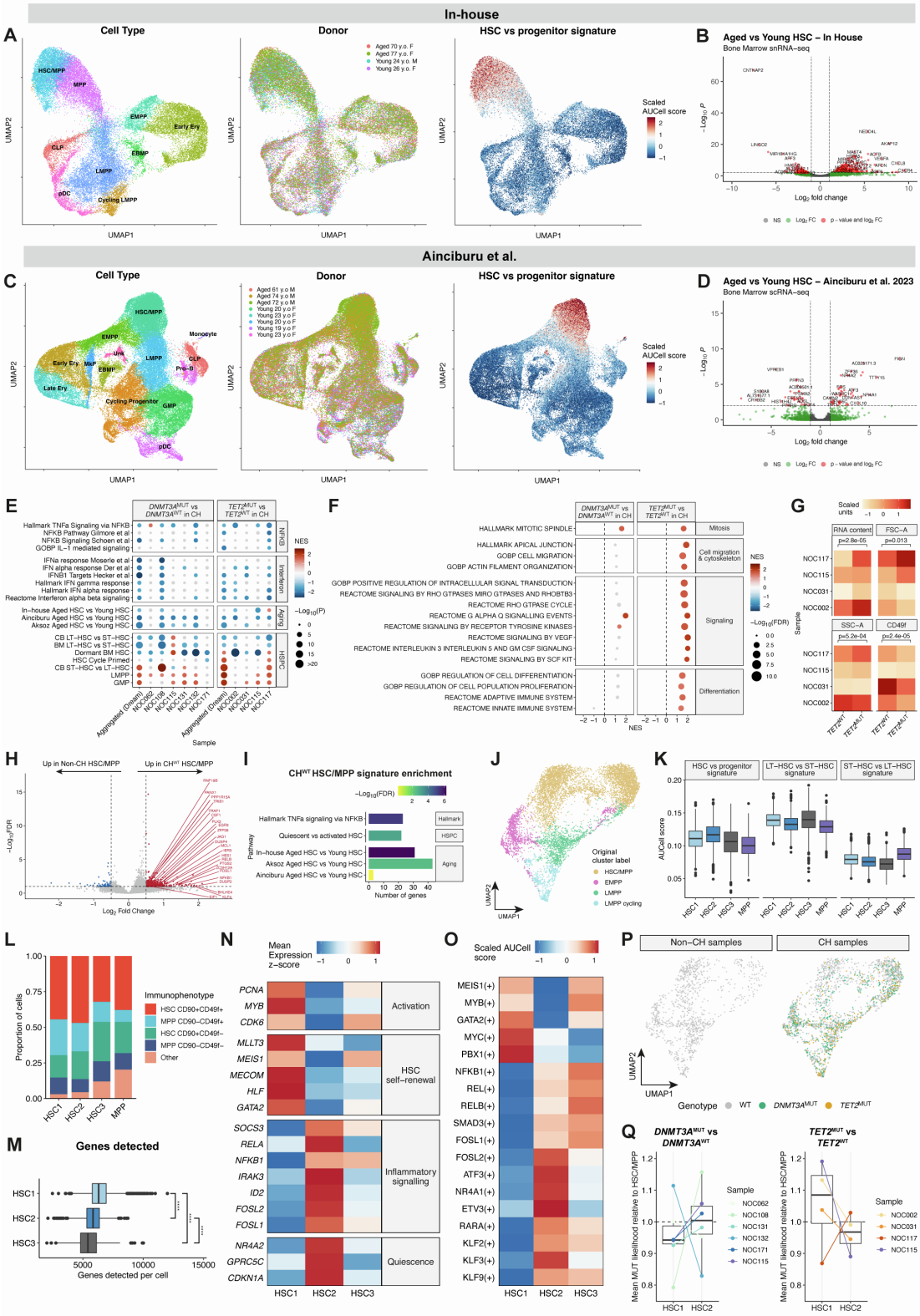


Figure S6. Generation of aged HSC gene signatures, consequences of CH mutations in

HSC/MPP, and characterization of HSC transcriptional heterogeneity, related to Figures 5 and 6.

(A) UMAPs showing snRNA-seq data for BM HSPCs from 2 young and 2 aged donors (in-house dataset) colored by Louvain cluster identity (left), donor identity (middle) and the scaled AUCell score for the HSC vs progenitor signature (right).

(B) Volcano plot showing differentially expressed genes (DEGs) between HSC/MPP from aged donors versus those from young donors from the in-house dataset. Genes with FDR < 0.01 and $\log_2$FC > 1 were considered differentially expressed (labelled red).

(C) UMAPs showing snRNA-seq data for BM HSPCs from 5 young and 3 aged donors (Ainciburu et al. dataset[13]) colored by Louvain cluster identity (left), donor identity (middle), and the scaled AUCell score for the HSC vs progenitor signature (right).

(D) Volcano plot showing DEGs between HSC/MPP from aged donors versus those from young donors from the Ainciburu et al[13] dataset. Genes with FDR < 0.01 and $\log_2$FC > 1 were considered differentially expressed (labelled red).

(E) Comparison of GSEA results between the aggregated Dream linear mixed model and differential expression analysis in individual samples using a Wilcox test. GSEA against NF-κB, interferon, and hematopoietic signatures comparing *DNMT3A*^{MUT} versus *DNMT3A*^{WT} HSC/MPPs and *TET2*^{MUT} versus *TET2*^{WT} HSC/MPPs within CH samples. Signatures with P-value > 0.05 are colored grey. Positive NES values indicate enrichment in mutant cells.

(F) GSEA against Hallmark, Gene Ontology biological process (GOBP), and Reactome signatures, comparing *DNMT3A*^{MUT} against *DNMT3A*^{WT} HSC/MPPs (left) and *TET2*^{MUT} against *TET2*^{WT} HSC/MPPs (right), within CH samples. Signatures with FDR > 0.2 are colored grey. Positive NES values indicate enrichment in mutant cells.

(G) RNA content (ratio between endogenous RNA reads and ERCC spike-in reads), cell size, granularity, and CD49f protein expression measured by FACS indexing, comparing *TET2*^{WT} against *TET2*^{MUT} HSC/MPPs. P-values calculated by linear mixed model.

(H) Volcano plot showing DEGs between CH^{WT} HSC/MPPs from 9 CH samples (n = 2622 cells) and WT HSC/MPPs from 4 non-CH samples (n = 1279 cells). Genes with FDR < 0.1 and $\log_2$FC > 0.5 were considered differentially expressed. Genes in the TNFα via NFKB pathway are labelled in red.

(I) Pathway enrichment of genes upregulated in CH^{WT} versus non-CH HSC/MPPs against Hallmark and hematopoietic signatures. Significance calculated by hypergeometric test.

(J) UMAP embedding of 8059 cells from the HSC/MPP, EMPP, LMPP and LMPP cycling clusters after feature weight derivation with the Self-Assembling Manifolds (SAM) algorithm. Cells are colored by the original cluster annotation used in previous figures.

(K) Comparison of AUCell expression scores for HSC gene signatures (indicated above each graph) in HSC and MPP subclusters (x-axis).

(L) Immunophenotype of cells in HSC subclusters (x-axis).

(M) Boxplot showing the number of genes detected per cell within the HSC1-3 clusters. ****p < 0.0001. Boxplots display the median and interquartile range. P-values calculated by unpaired *t*-test.

(N) Expression of genes related to HSC quiescence, activation, self-renewal, and inflammatory signaling within the HSC1-3 clusters.

(O) Heatmap showing scaled AUCell scores for selected regulons that are differentially active in the HSC1-3 clusters.

(P) UMAP embeddings showing cells from non-CH and CH samples. Cells are colored by genotype. For clarity, only WT and single-mutant cells are shown.

(Q) Boxplots showing the distribution of mean mutant clone likelihood in the HSC1 and HSC2 clusters, within *DNMT3A*^{MUT} and *TET2*^{MUT} CH samples. Boxplots display the median and interquartile range. Each dot represents a CH sample (color legend on the right). Mutant clone likelihood is normalized to the mean within the HSC/MPP clusters.

Supplemental References

1. Hecker, J.S., Hartmann, L., Rivière, J., Buck, M.C., van der Garde, M., Rothenberg-Thurley, M., Fischer, L., Winter, S., Ksienzyk, B., Ziemann, F., et al. (2021). CHIP and hips: clonal hematopoiesis is common in patients undergoing hip arthroplasty and is associated with autoimmune disease. Blood *138*, 1727-1732. https://doi.org/10.1182/blood.2020010163.

2. Abelson, S., Collord, G., Ng, S.W.K., Weissbrod, O., Mendelson Cohen, N., Niemeyer, E., Barda, N., Zuzarte, P.C., Heisler, L., Sundaravadanam, Y., et al. (2018). Prediction of acute myeloid leukaemia risk in healthy individuals. Nature *559*, 400-404. https://doi.org/10.1038/s41586-018-0317-6.

3. van Zeventer, I.A., de Graaf, A.O., Salzbrunn, J.B., Nolte, I.M., Kamphuis, P., Dinmohamed, A., van der Reijden, B.A., Schuringa, J.J., Jansen, J.H., and Huls, G. (2023). Evolutionary landscape of clonal hematopoiesis in 3,359 individuals from the general population. Cancer Cell. https://doi.org/10.1016/j.ccell.2023.04.006.

4. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nature Cell Biology *19*, 271-281. https://doi.org/10.1038/ncb3493.

5. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. Cell *144*, 296-309. https://doi.org/10.1016/j.cell.2011.01.004.

6. Drissen, R., Thongjuea, S., Theilgaard-Mönch, K., and Nerlov, C. (2019). Identification of two distinct pathways of human myelopoiesis. Science Immunology *4*, eaau7148-eaau7148. https://doi.org/10.1126/sciimmunol.aau7148.

7. Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., Meira, A.R., et al. (2021). Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. Cell Reports *36*, 109698-109698. https://doi.org/10.1016/j.celrep.2021.109698.

8. Xie, S.Z., Kaufmann, K.B., Wang, W., Chan-Seng-Yue, M., Gan, O.I., Laurenti, E., Garcia-Prat, L., Takayanagi, S.-i., Ng, S.W.K., Xu, C., et al. (2021). Sphingosine-1-Phosphate Receptor 3 Potentiates Inflammatory Programs in Normal and Leukemia Stem Cells to Promote Differentiation. Blood Cancer Discovery *2*, 32-53. https://doi.org/10.1158/2643-3230.BCD-20-0155.

9. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., et al. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nature Immunology *19*, 85-97. https://doi.org/10.1038/s41590-017-0001-2.

10. Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., et al. (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science *345*, 1251033-1251033. https://doi.org/10.1126/science.1251033.

11. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189-196. https://doi.org/10.1126/science.aad0501.

12. Zhang, Y.W., Mess, J., Aizarani, N., Mishra, P., Johnson, C., Romero-Mulero, M.C., Rettkowski, J., Schönberger, K., Obier, N., Jäcklein, K., et al. (2022). Hyaluronic acid–GPRC5C signalling promotes dormancy in haematopoietic stem cells. Nature Cell Biology *24*, 1038-1048. https://doi.org/10.1038/s41556-022-00931-x.

13. Ainciburu, M., Ezponda, T., Berastegui, N., Alfonso-Pierola, A., Vilas-Zornoza, A., San Martin-Uriz, P., Alignani, D., Lamo-Espinosa, J., San-Julian, M., Jiménez-Solas, T., et al. (2023). Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. eLife *12*, e79363. https://doi.org/10.7554/eLife.79363.

A.3 GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells

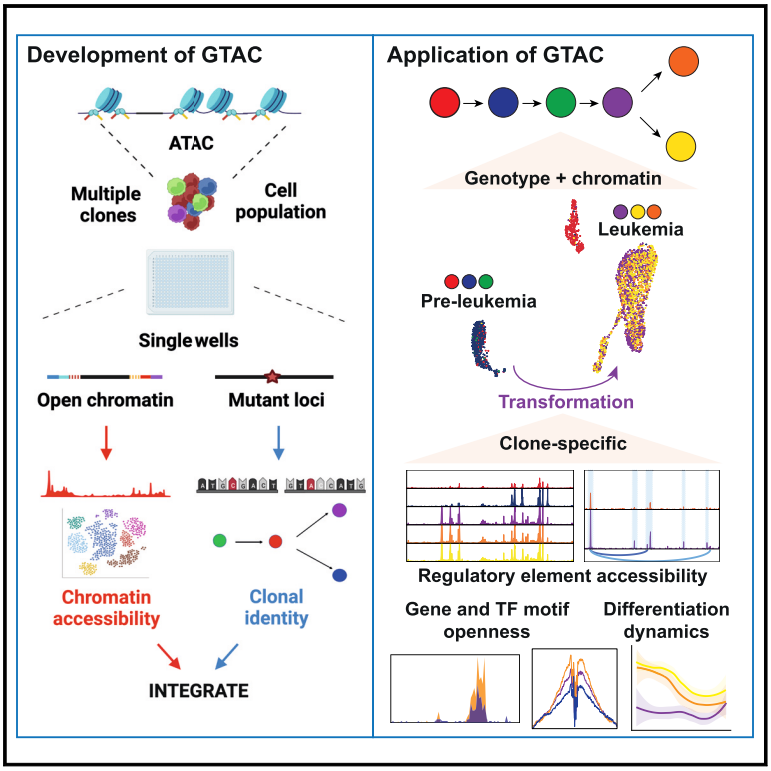
pdfpages

Cell Stem Cell

Resource

GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells

Graphical abstract



Highlights

- High-sensitivity genomic genotyping linked to scATAC-seq in single malignant cells
- Characterization of epigenetic variation as preleukemic clones transform to leukemia
- Differentiation of specific leukemic clones is delayed at an LSC-like chromatin state
- Genomic elements perturbed in LSC-like clones relate to stemness and differentiation

Authors

Sven Turkalj, Niels Asger Jakobsen, Angus Groom, ..., James O.J. Davies, Jim R. Hughes, Paresh Vyas

Correspondence

paresh.vyas@imm.ox.ac.uk

In brief

Turkalj, Jakobsen, et al. develop a method linking clonal identities of single cells to their chromatin accessibility and study how epigenetic landscapes change with clonal evolution in human acute myeloid leukemia, revealing altered accessibility of genes and transcription factor motifs linked to leukemia stemness downstream of cooperating somatic mutations.

Turkalj et al., 2023, Cell Stem Cell *30*, 722–740
May 4, 2023 © 2023 Published by Elsevier Inc.
https://doi.org/10.1016/j.stem.2023.04.012

CellPress

CellPress

Cell Stem Cell

Resource

GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells

Sven Turkalj,[1,2,6] Niels Asger Jakobsen,[1,2,3,6] Angus Groom,[1,2] Marlen Metzner,[1,2] Simone G. Riva,[1,4] E. Ravza Gür,[1,4] Batchimeg Usukhbayar,[1,2] Mirian Angulo Salazar,[1,2] Lance D. Hentges,[1,4] Gerda Mickute,[1,2] Kevin Clark,[5] Paul Sopp,[5] James O.J. Davies,[1,2,3] Jim R. Hughes,[1,4] and Paresh Vyas[1,2,3,7,*]

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[2]Oxford Centre for Haematology, NIHR Oxford Biomedical Research Centre, Oxford, UK
[3]Department of Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[4]MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[5]Flow Cytometry Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[6]These authors contributed equally
[7]Lead contact
*Correspondence: paresh.vyas@imm.ox.ac.uk
https://doi.org/10.1016/j.stem.2023.04.012

SUMMARY

Understanding clonal evolution and cancer development requires experimental approaches for characterizing the consequences of somatic mutations on gene regulation. However, no methods currently exist that efficiently link high-content chromatin accessibility with high-confidence genotyping in single cells. To address this, we developed Genotyping with the Assay for Transposase-Accessible Chromatin (GTAC), enabling accurate mutation detection at multiple amplified loci, coupled with robust chromatin accessibility readout. We applied GTAC to primary acute myeloid leukemia, obtaining high-quality chromatin accessibility profiles and clonal identities for multiple mutations in 88% of cells. We traced chromatin variation throughout clonal evolution, showing the restriction of different clones to distinct differentiation stages. Furthermore, we identified switches in transcription factor motif accessibility associated with a specific combination of driver mutations, which biased transformed progenitors toward a leukemia stem cell-like chromatin state. GTAC is a powerful tool to study clonal heterogeneity across a wide spectrum of pre-malignant and neoplastic conditions.

INTRODUCTION

Somatic genetic variation leads to clonal mosaicism,[1] pre-cancerous clonal outgrowth,[2,3] and neoplastic transformation.[4,5] Cancerous clonal evolution driven by serial and/or branched acquisition of somatic mutations is a cause of intra-tumoral heterogeneity,[6] therapy resistance,[7] and disease relapse.[8,9] Hence, knowledge of how somatic mutations corrupt cellular phenotype is a prerequisite for understanding the mechanisms underlying clonal dynamics and for the development of personalized therapeutic approaches.

Chromatin is the medium through which transcription factors (TFs), epigenetic regulators, and signaling pathways control gene expression and cell phenotype.[10,11] Epigenetic dysregulation is a common feature in a wide range of human cancers, contributing to disease progression through activation of oncogenes, silencing of tumor suppressor genes, or disruption of differentiation programs.[12] Furthermore, many human cancers and pre-cancerous clonal outgrowths harbor somatic mutations in genes encoding epigenetic regulators, which have a direct effect on gene regulation, clonal expansion, and differentiation bias.[2,12–15] An example of such cancers are hematopoietic neoplasms such as acute myeloid leukemia (AML).[16] Methods that enable the epigenetic consequences of somatic variation to be characterized in primary patient samples are therefore critical to understand the mechanisms underlying cancer development.

Recent strategies combining single-cell RNA sequencing (scRNA-seq) and targeted genotyping allow linkage of genetic lineage history with transcriptomic heterogeneity within primary human samples, in which mutant and wild-type (WT) cells are often admixed without any known surface markers allowing for their purification.[14,17–20] Albeit powerful, these approaches fail to capture epigenetic variation and, specifically, chromatin accessibility, limiting our understanding of how somatic variation is linked to epigenetic state and TF activity. The assay for transposase-accessible chromatin with sequencing (ATAC-seq)[10,21,22] and single-cell ATAC-seq (scATAC-seq)[23–25] have enabled the study of chromatin accessibility variation across heterogeneous cell types while methods coupling scATAC-seq with transcriptome sequencing[26–29] and surface[30] or intracellular protein profiling[31] capture multiple layers of molecular and phenotypic information. Finally, integration of mitochondrial

genome sequencing with chromatin accessibility linked clonal dynamics with epigenetic state.[32] However, no method links chromatin with driver somatic events at high efficiency in primary tissues, representing a major unmet need in our analyses of somatic disease.

Here, we present a technique that addresses this methodological gap, termed Genotyping with the Assay for Transposase-Accessible Chromatin (GTAC), which couples single-cell genotyping of multiple genomic loci with robust scATAC-seq in FACS-sorted nuclei. We first benchmarked GTAC in cell lines, showing accurate mutation calling enabled by parallel multi-allelic capture of several genomic targets while maintaining a high-quality scATAC library. Next, by applying GTAC to a primary AML sample, we captured clone-specific chromatin accessibility at defined stages of differentiation, providing mechanistic insights into the epigenetic consequences of specific driver mutations within a complex clonal hierarchy.

RESULTS

GTAC workflow

Coding mutations are not consistently covered in scATAC libraries because most known driver mutations occur within exons, whereas fragments from scATAC libraries map mostly to regulatory genomic loci. We therefore designed GTAC, a method for simultaneous genotyping of multiple driver mutations and high-quality chromatin accessibility profiling from the same cell (Figure 1). Cell lysis and Tn5 tagmentation are performed on a cell population, followed by FACS of single nuclei into plates.[33,34] Genotyping targets are captured by the addition of target-specific primers to the initial PCR, which amplifies genomic targets in parallel with fragments of accessible chromatin. Careful optimization of primer design (STAR Methods) is crucial for minimal interference with scATAC library generation. We optimized both the PCR protocol and the indexing strategy to improve scATAC library complexity and increase cell throughput. The incorporation of 384 "i7" cell barcodes and "i5" plate barcodes during the initial PCR step enables scATAC libraries from many plates to be sequenced together without additional PCR reactions. After the PCR, 80% of the material is pooled and purified to yield scATAC libraries. The remaining 20% of the initial library material is transferred to another plate and used for further target amplification via two additional PCR reactions (PCR1 and PCR2), which fully barcode the genotyping libraries. scATAC and genotyping libraries are then sequenced and pre-processed separately. After single-cell mutation calling for each amplified locus, genotype information for every cell is linked to its accessibility profile via cell barcodes (STAR Methods).

GTAC allows for sensitive mutation detection at multiple genomic loci in cell lines

To assess efficient capture of genotyping targets, we first performed GTAC on K562 cells, amplifying previously reported mutant or variant loci (Figure 2A; Table S1). We compared coverage across the 5 loci in K562 cells in libraries prepared with standard scATAC without genotyping primers and GTAC (Figure 2B). The median number of reads covering the variant nucleotide was 0 for scATAC libraries, whereas it was 7,433 reads for GTAC libraries across all loci, indicating that genotyping primers are needed for efficient locus capture. Target amplification was equally efficient in the absence of Tn5 (Figure 2C), showing that locus capture is unrelated to Tn5 integration. This indicates that genotyping locus amplification does not depend on its chromatin accessibility.

We next benchmarked the accuracy and sensitivity of genotyping, targeting previously reported mutant or variant loci in several cell lines, using THP1 cells as negative WT control (Figure 2D). For all 5 mutant or variant loci in K562, the single-cell variant allele frequency (scVAF) of mutant reads was <1% in THP1 cells (Figures 2E and S1A–S1E), allowing us to set thresholds for mutation calling. scVAF distributions observed in K562 were consistent with the reported zygosity[35] (Figure 2E; note that K562 is a triploid cell line in which a heterozygous mutation is present on either 1 out of 3 or 2 out of 3 alleles). In addition, we targeted *RUNX1* and *NRAS* loci in JURKAT and OCI-AML3 cells, respectively, and observed a marked difference in scVAF distributions comparing the two (Figure 2F), consistent with the *RUNX1* mutation being heterozygous in JURKAT (average scVAF was 0.47) and the *NRAS* mutation being homozygous in OCI-AML3 cells (average scVAF was 0.99),[35] providing evidence that our genotyping data can discriminate between a heterozygous and a homozygous mutant population.

Next, we estimated allelic dropout (ADO) rates for the profiled heterozygous loci across cell lines (Figure 2G). ADO, caused by failed amplification of one allele during PCR, can result in misassignment of heterozygous mutant loci as WT, compromising the accuracy of downstream analysis. For the nine heterozygous loci, we considered that ADO occurred if a heterozygous mutant cell line was wrongly assigned as homozygous mutant or homozygous WT. By using scVAF values from THP1 cells to set thresholds for genotype calling (Figures S1A–S1J), we observed variable but consistently low ADO rates across the profiled loci, ranging from 1.3% for *NOTCH1* to 11.9% for *HOXA9* (Figure 2G).

Finally, given that accurate genotyping of multiple loci is crucial for high-confidence reconstruction of clonal hierarchies, we tested whether GTAC could reliably detect multiple mutations in the same single cells by simultaneously amplifying 5 mutant loci in K562 cells. All 5 mutations were detected in 79% of cells, at least 4 mutations were detected in 94%, while only 6% of cells failed genotyping at 2 or more loci (Figure 2H). Given that failed genotyping of an early mutation can still lead to correct clonal assignment if later mutations are efficiently detected in the same cells (STAR Methods), this suggests that GTAC can be used for correct clonal assignment of most single cells. We thus demonstrate that GTAC enables high-confidence, accurate genotyping of multiple genomic loci within single cells.

Targeting multiple genotyping loci does not disrupt scATAC library quality

We next assessed whether GTAC consistently yielded high-quality scATAC libraries with increasing numbers of genotyped loci. We performed GTAC on K562 cells targeting 4, 6, or 8 loci and compared scATAC library metrics with those of a scATAC-only control. All genotyping amplicons were detected in over 90% of cells in all GTAC conditions (98.9%, 92.6%, and 93.7%, respectively), indicating no interference between genotyping amplicons when increasing numbers of loci are genotyped (Figure S1K).

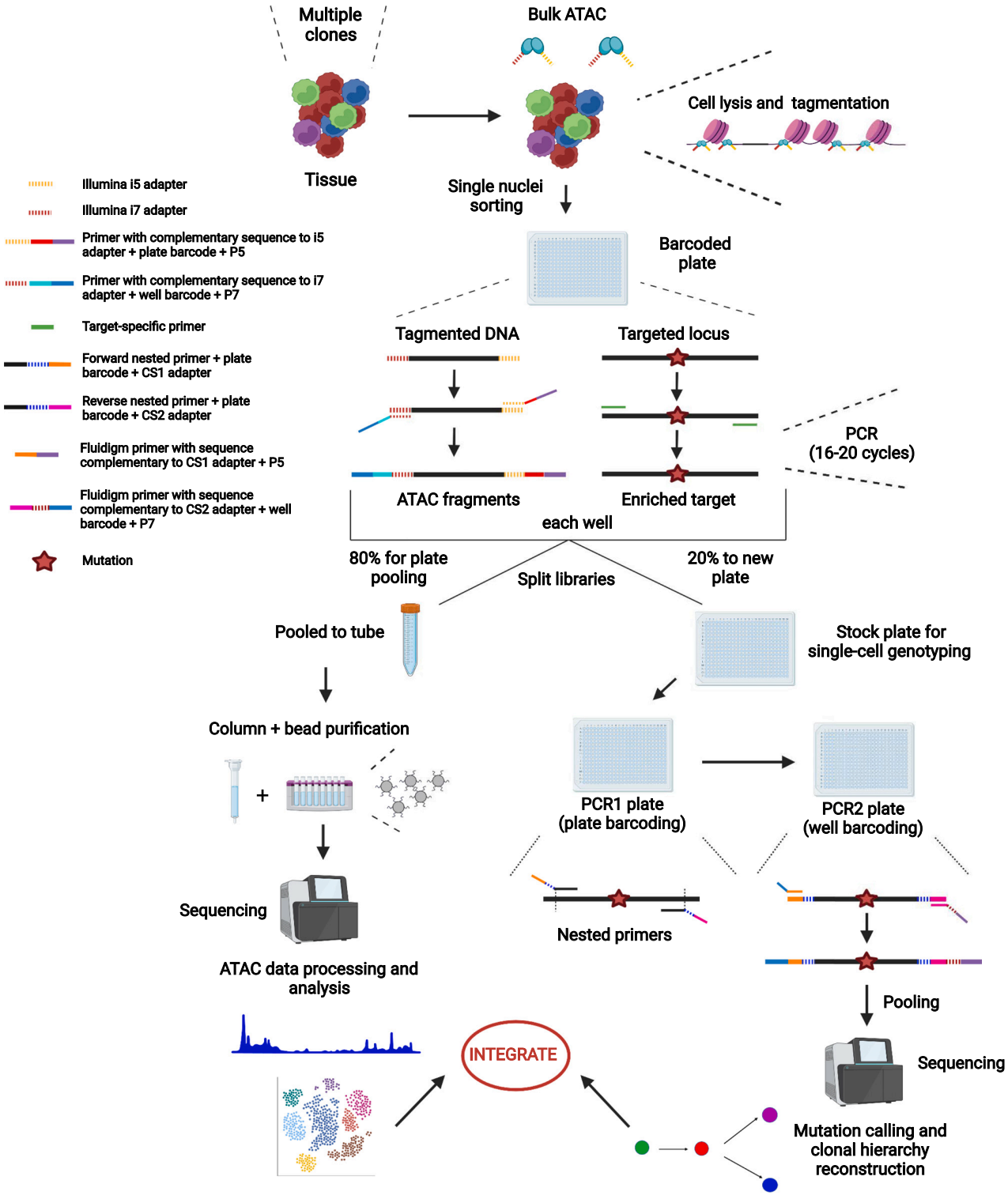


Figure 1. GTAC workflow

Schematic overview of GTAC. Primer legend is shown on the left. Details regarding experimental procedures are described in STAR Methods.

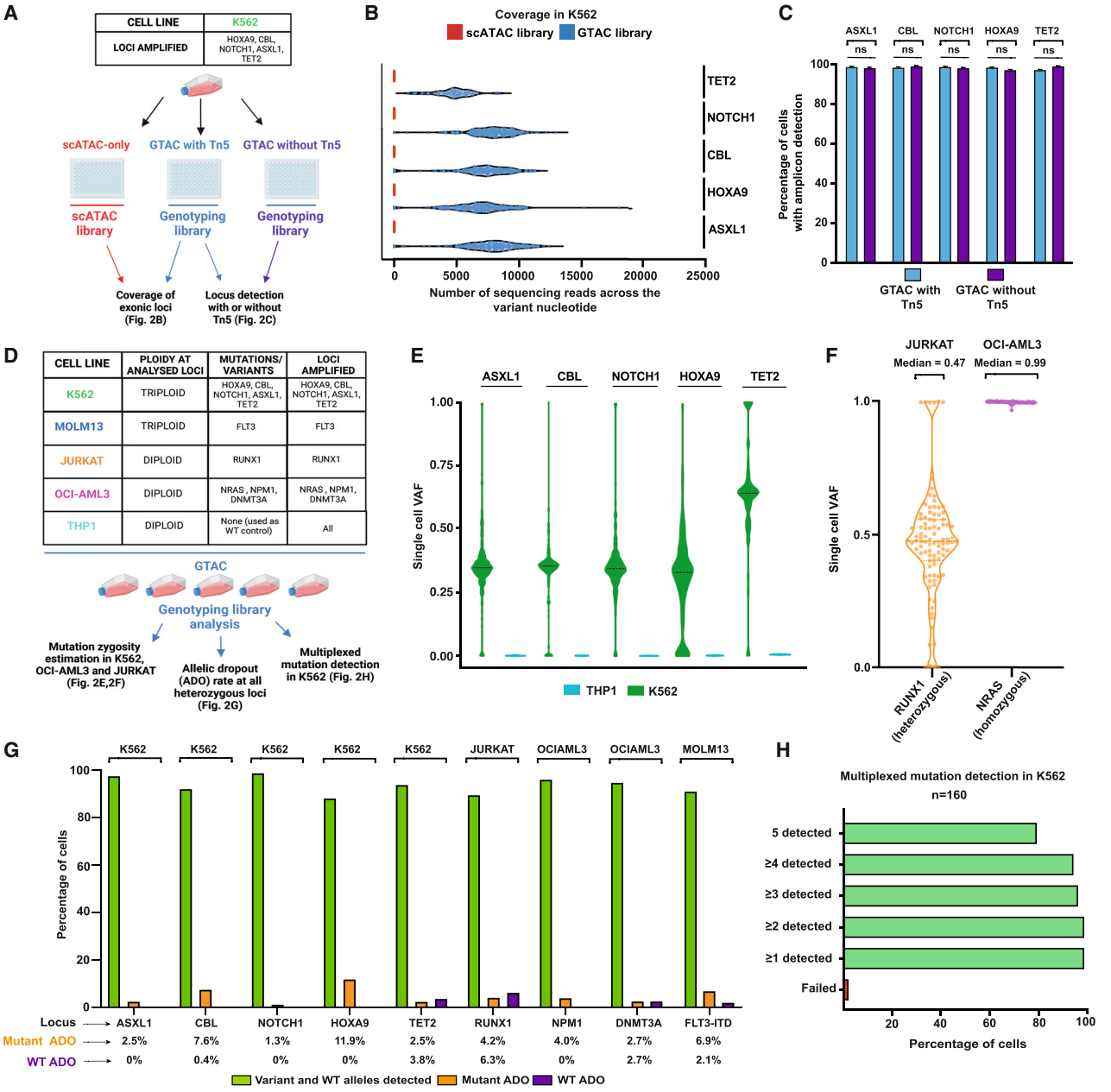


Figure 2. GTAC enables sensitive mutation detection from multiple amplified genomic loci

(A) Workflow for validation of genotyping amplicon capture in K562 cells, using scATAC alone (red) or the GTAC assay with (blue) or without (purple) the Tn5 enzyme.

(B) Number of sequencing reads in single K562 cells across the variant nucleotide in the indicated loci, comparing scATAC only (red) and GTAC libraries (blue). Median coverage in scATAC and GTAC libraries, respectively, for each locus: *TET2*: 0 and 4,823; *NOTCH1*: 0 and 7,993; *CBL*: 0 and 7,461; *HOXA9*: 0 and 6,961; *ASXL1*: 0 and 7,980. Number of cells tested by GTAC is 240 for *ASXL1*, *CBL*, *HOXA9*, and *NOTCH1*; 80 for *TET2*. Number of cells tested by scATAC is 95 for each locus.

(C) Percentage of K562 cells showing ≥30 sequencing reads across the amplified loci, comparing conditions with (blue) or without (purple) Tn5. Error bars show the standard error. Fisher's exact test was used (ns, not significant).

(D) Single-cell genotype calling validation workflow. Top: table shows cell lines tested, ploidy at interrogated loci, mutations/variants expected, and genomic loci amplified. Bottom: cells were tested with GTAC, and mutation calling was performed.

(E) Single-cell VAF (scVAF) values for variant loci in K562, in cells with ≥30 reads across the variant nucleotide. Median scVAF in K562 and THP1 cells (WT control), respectively, are: *ASXL1*: 0.345 and 0.001; *CBL*: 0.352 and 0.002; *NOTCH1*: 0.342 and 0.001; *HOXA9*: 0.327 and 0.002; *TET2*: 0.642 and 0.001. Median values are shown.

(F) scVAF for the *RUNX1* locus in JURKAT cells and for the *NRAS* locus in OCI-AML3 cells. Each dot is a single cell (n = 95, each). Median scVAF for *RUNX1* and *NRAS* is 0.471 and 0.996, respectively.

(legend continued on next page)

Library complexity (Figure 3A) and signal-to-noise ratio (fraction of reads in peaks; FRIP) (Figure 3B) were comparable across GTAC conditions and not significantly different from the scATAC control. GTAC showed improved library complexity and FRIP compared with 10X Chromium droplet-based technology.[36] Median mitochondrial read contamination was <1% in all GTAC and our scATAC conditions, in contrast to >5% with the published plate-based method (Figure 3C). Additionally, transcription start site (TSS) enrichment (Figure 3D) and aggregated fragment size distribution (Figure 3E) were unaltered between GTAC and scATAC. While most cells displayed TSS enrichment values above quality control (QC) thresholds, subtly lower median values in our conditions compared with prior data (Figure 3D) are likely related to more efficient capture of di-nucleosomal fragments (Figure S1L; eukaryotic promoters are largely nucleosome-free[37]). Comparison of peak calls showed strong correlation between the three GTAC and the scATAC-seq conditions (r = 0.96 for all comparisons) and with two public K562 datasets produced via plate-based scATAC[33] (Figure 3F). In addition, visual inspection of aggregated data from single cells suggested no qualitative differences between these conditions (Figure 3G). Importantly, visual inspection of scATAC genomic coverage surrounding the profiled mutant loci in K562 cells showed no difference between conditions, suggesting no disruption of scATAC fragment amplification in proximity of genotyped loci (Figure S1M). We observed >90% of cells passing scATAC QC regardless of the number of genotyped loci, comparable to standard plate-based scATAC[33] (Figure 3H). Likewise, the percentage of cells passing scATAC QC in which more than 30 sequencing reads were obtained for all genotyped loci was >90% in all conditions (Figure 3I). Crucially, the percentage of cells passing scATAC QC in which all profiled heterozygous mutations were correctly detected was 80% or higher across conditions (Figure 3J), leading to high numbers of cells useful for analysis.

To test integration of scATAC and genotyping data, we performed a cell line mixing experiment with K562, JURKAT, and OCI-AML3 cells, targeting 3 genomic loci in each cell line (1 mutant and 2 WT loci per cell line). Clustering based on scATAC data alone revealed three distinct clusters (Figure 3K). Integrating chromatin accessibility data with genotyping information revealed correct mutational calls for 97.2% of cells (Figures 3L–3N). Taken together, we demonstrated that GTAC generates high-quality scATAC data unperturbed by targeted genotyping of multiple loci and that scATAC and genotyping data can be integrated.

GTAC links clonal hierarchy to accessibility landscape and identifies restriction of clones to distinct differentiation stages in human acute myeloid leukemia

To investigate how clonal structure correlates with chromatin accessibility in a primary malignant tissue, we applied GTAC to 3,520 AML cells from a patient with expanded granulocyte-monocyte progenitor (GMP) and lymphoid-primed multipotent progenitor (LMPP) immunophenotypic compartments (Figures S2A and S2B). We genotyped 5 mutant loci previously identified by bulk sequencing (Figure 4A), as well as two heterozygous germline single-nucleotide polymorphisms (SNPs), to estimate ADO in primary sample (Figure S2C). Distribution of scVAFs for two germline SNPs in *SH2B3* and *ASXL2* indicated very low WT (2.0% and 3.4%) and variant (2.2% and 3.1%) ADO rates (Figure S2D). This allowed biallelic detection in 95.8% of cells for the *SH2B3* locus and in 93.5% of cells for the *ASXL2* locus, which was consistent across plates (Figure S2E).

To reconstruct clonal hierarchy, we performed mutation calling on the 5 mutant loci. High genotyping success rates (from 94% to 98% per amplicon) resulted in 3,409 cells (96.8%) being assigned to a clone (Figure 4B). Moreover, 3,113 cells (88.44%) passed stringent scATAC QC (STAR Methods) and were successfully genotyped (Figure 4C). Mutant cell scVAF distributions revealed that *TET2* and *ASXL1* mutations were heterozygous, while *RUNX1* and the two X-linked *BCOR* mutations were homozygous and hemizygous, respectively (Figure 4D). Of note, the *RUNX1* R201G missense mutation disrupts DNA binding,[38–40] suggesting that *RUNX1* R201G homozygous mutant cells in this sample harbor a complete *RUNX1* loss of function (LoF). By integrating genotype information across loci (Figure 4E), we reconstructed the most likely clonal hierarchy in which *TET2*, *ASXL1*, and *RUNX1* mutations were acquired in linear fashion (T, TA, and TAR clones), followed by branched acquisition of two distinct *BCOR* mutant clones (TARB1 and TARB2) (Figure 4F). Notably, mutant cell fractions obtained with GTAC correlated well with clone size estimates from bulk sequencing (Figure S2F).

scATAC data analysis indicated high library quality (Figures S2G–S2K), with improved metrics as compared with those obtained for leukemia samples processed with commercial droplet-based technology[41] (Figures S2G–S2I). We first used chromatin accessibility data to perform dimensionality reduction and clustering agnostic to genotyping information. Of note, we did not observe plate-related batch effects (Figure S2L). Cluster identities (Figure 4G) were assigned using gene activity (i.e., accessibility) scores of key marker genes, binding motif accessibility for key transcription factors, and differential enrichment of marker peaks (Figures S3A–S3D; STAR Methods). Projection of our dataset onto a healthy human hematopoiesis scATAC-seq[41] (Figure S3E; STAR Methods) was broadly consistent with our cluster assignment, though there were differences, which is unsurprising given that leukemogenic mutations disrupt differentiation trajectories. Readout of chromatin openness in regulatory and/or coding regions confirmed higher accessibility of the quiescence-related *PBX1* gene[42] in the hematopoietic stem cells/multipotent progenitors (HSCs/MPPs) and megakaryocytic/erythroid progenitors (Mk/EPs), of *GATA1* in Mk/EP cells, of *CD3E/D/G*

(G) Percentage of cells in which both mutant and WT (green), only the mutant (purple), or only the WT (orange) allele was detected, across 9 heterozygous loci. Cell lines are indicated above. n = 236 (*ASXL1*), 238 (*CBL*), 235 (*HOXA9*), 232 (*NOTCH1*), 80 (*TET2*), 93 (*RUNX1*), 149 (*NPM1*), 151 (*DNMT3A*), and 144 (*FLT3*). Mutant and WT allelic dropout (ADO) rates for each locus are indicated below.

(H) Frequency of detection of heterozygous mutations in K562 cells when 5 mutant loci were amplified in parallel. The percentage of cells in which all 5 (5 detected; 79%) or at least 4 (≥4 detected; 94%), 3 (≥3 detected; 96%), 2 (≥2 detected; 98%), or 1 (≥1 detected; 98%) mutant loci were detected is shown. 2% of cells failed genotyping at all loci (Failed).

CellPress

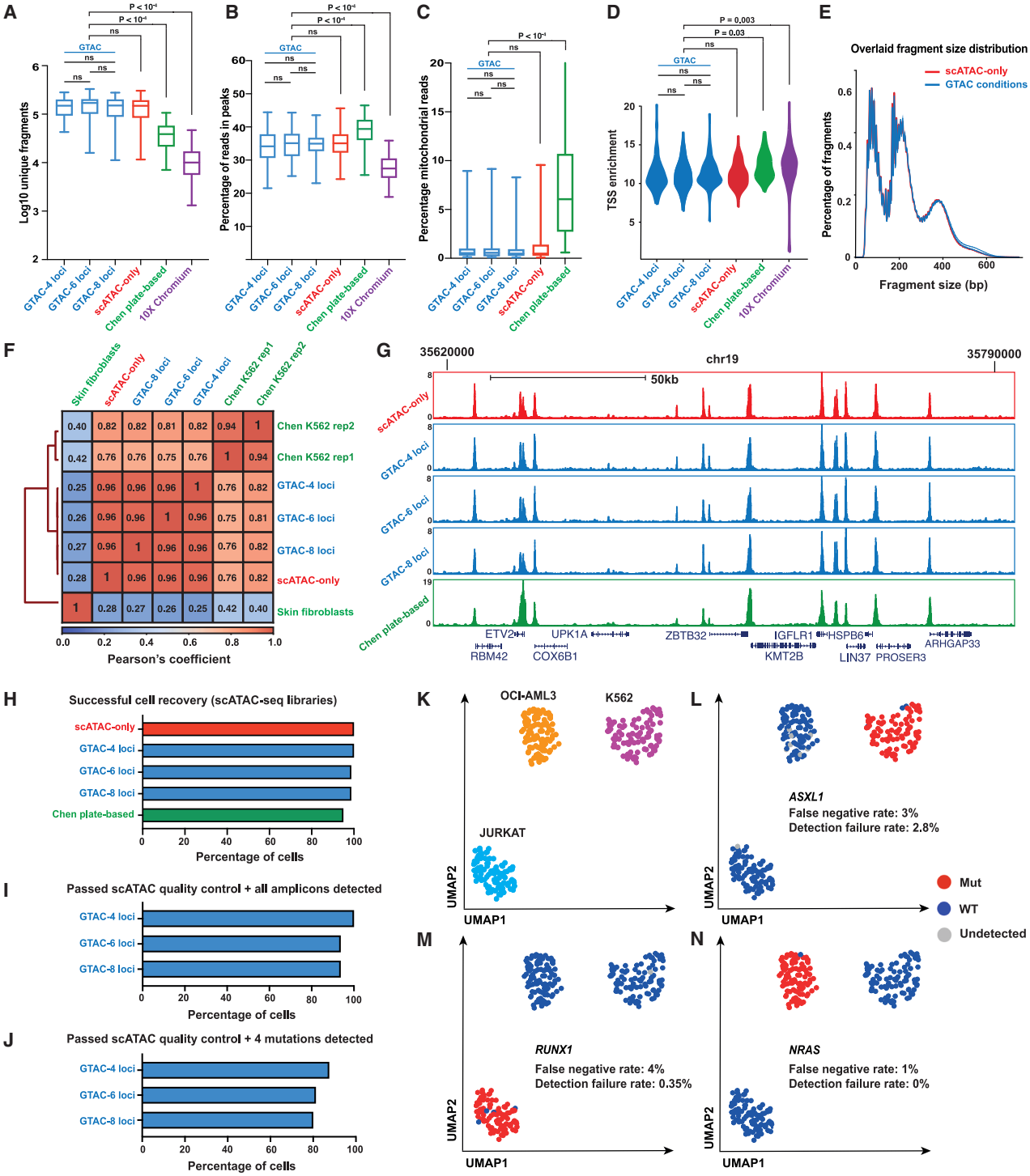


Figure 3. scATAC library quality is not altered with increasing numbers of amplified genomic loci

(A–H) 80 K562 cells were assayed in 4 different conditions: GTAC (blue) where either 4, 6, or 8 genotyping loci were co-amplified or scATAC alone (red). In (A)–(D), (F), and (H), metrics were compared with published plate-based K562 scATAC-seq data[33] (green, Chen plate-based); in (A), (B), and (D), metrics were compared with published K562 data obtained with a 10X Genomics droplet-based method[36] (purple, 10X Chromium). In (A)–(D), ordinary one-way ANOVA with Dunnett's multiple comparisons was used (ns, non-significant, adj. $p > 0.05$).

(A) Log_{10} of the number of unique nuclear ATAC fragments. Median values are: GTAC with 4 loci: 5.172; GTAC with 6 loci: 5.231; GTAC with 8 loci: 5.178; scATAC alone: 5.174; Chen plate-based: 4.589; 10X Chromium: 4.001. Median values are shown within boxplots. Whiskers extend from the 2.5th to the 97.5th percentile.

(legend continued on next page)

genes in T/NK cells, and of *CD79B* in the B progenitor-like cluster (Figure S4A). In contrast, the *SETBP1* gene, preferentially expressed in leukemic stem cells (LSCs),[43] and *FLT3* were more accessible in the clusters resembling lympho-myeloid progenitors (Figure S4B).

Next, by projecting AML-driver mutation clonal structure onto chromatin accessibility data, we noticed a striking segregation of genotypes into separate epigenomic entities (Figure 4H). 97.5% of T/NK cells were WT with the remaining cells being *TET2* mutant. 87.1% of HSC/MPP and 82.5% of Mk/EP clusters were composed of *TET2* single-mutant cells intermingled with a minority of *TET2/ASXL1* double-mutant cells, suggesting that these populations were preleukemic.[3,44,45] In stark contrast, *TET2/ASXL1* double-mutant clones that acquired either a *RUNX1* mutation or *RUNX1* and *BCOR* mutations were distributed almost exclusively within the lympho-myeloid, myeloid, and B cell progenitor-like clusters, contributing to more than 99% of these populations. We and others have previously shown that the AML LSC populations reside in these populations.[46–48] This suggests that in this patient, the *RUNX1* mutation acts as a key transforming event, leading from preleukemia to leukemia. Notably, *RUNX1* mutant clones displayed increased accessibility of several genes reported as upregulated in *RUNX1* mutant compared with *RUNX1* WT AML,[49,50] some of which are lymphoid markers, *DNTT*, *BLNK*, *IRF8*, and *MS4A1,* as well as *MN1,* associated with worse AML prognosis (Figure S4C).[51] Although *BCOR* mutant clones did not form separate progenitor clusters within the leukemic progenitor populations, they were significantly enriched within the most primitive LMPP-like differentiation stage (Figure 4I).

Unsurprisingly, TF motif openness (Figure 4J) reflected the differentiation stage of the clones. WT cells were enriched for DNA-binding motifs for T cell-associated TFs EOMES,[52,53] LEF1,[54] and RUNX.[55] The preleukemic T/TA clones were enriched for TF motifs important in HSCs/MPPs (HOX, GATA2) and in megakaryocytic and erythroid cells (GATA1, the heterodimeric bZIP TF NFE2/MAFF, and members of the nuclear factor family NFIA/NFIC). Finally, in the leukemic TAR/TARB1/TARB2 clones, enriched TF motifs included those important for lympho-myeloid progenitor differentiation[56,57] (SPI1, CEBP, and members of the IRF, AP-1, MEF2, and TCF families). MEF2C/D factors, whose motifs were particularly enriched in *BCOR* mutants, have been implicated in AML chemotherapy resistance,[58–60] and MEF2C is upregulated in LSCs compared with leukemic progenitor cells.[43] Footprinting analyses confirmed higher DNA binding of the EOMES in T/NK cells, HOXB4 in preleukemic clones, and SPI1 in leukemic clones (Figures S4D).

To understand whether the segregation of leukemic and preleukemic clones into different chromatin states was mainly driven by the *RUNX1* mutation or by a difference in differentiation stage, we integrated genotyping information with the healthy human hematopoiesis reference projection[41] (Figure S4E). This showed that 97 of 595 genetically preleukemic (*T* and *TA*) cells, as well as most genetically leukemic (TAR/TARB1/TARB2) LMPP-like cells, mapped to the normal CMP/LMPP population (Figures S3E, S4E, and S4F). In contrast, only 1 of 2,141 genetically leukemic (TARB2) cells mapped to healthy HSC. This suggests two possibilities: (1) the preleukemic cell of origin of leukemic clones transited through the CMP/LMPP-like stage when *RUNX1* mutation was acquired, or (2) a preleukemic HSC acquired the *RUNX1* mutation, but this caused clonal dominance of its progeny at the LMPP stage. Importantly, both scenarios suggest profound chromatin reprogramming at the LMPP stage upon *RUNX1* mutation acquisition on the background of preleukemic mutations and are concordant with our prior finding that the LMPP population is the closest normal counterpart of LMPP-like LSCs.[46] Finally, to understand chromatin changes associated specifically with transformation, we compared the accessibility profiles of genetic preleukemic and leukemic cells that mapped to the normal CMP/LMPP compartment. Interestingly, while *RUNX1* gene activity was greater in leukemic CMP/LMPP cells (Figure S4G), its binding motif was less accessible (Figure S4H), consistent with disrupted DNA binding. Leukemic cells mapping to CMP/LMPP also showed higher accessibility of WT1 and MEIS1 motifs (Figure S4H), linked to AML maintenance.[61,62] Consistent with higher MYC motif accessibility in leukemic cells mapping to CMP/LMPP (Figure S4H), accessible genes were enriched for pathways associated with proliferation (MYC targets and cell cycle regulators), MAPK, and PI3K signaling (Figure S4I). These data, which take differentiation

(B) Percentage of ATAC reads in called peaks. Median values are: GTAC with 4 loci: 34.2%; GTAC with 6 loci: 35.1%; GTAC with 8 loci: 35.0%; scATAC alone: 35.0%; Chen plate-based: 39.4%; 10X Chromium: 27.5%. The rest is as in (A).
(C) Percentage of reads mapping to the mitochondrial genome. Median values are: GTAC with 4 loci: 0.53%; GTAC with 6 loci: 0.56%; GTAC with 8 loci: 0.47%; scATAC alone: 0.48%; Chen plate-based: 6.05%. The rest is as in (A).
(D) Transcription start site (TSS) enrichment. Median values are: GTAC with 4 loci: 10.65; GTAC with 6 loci: 10.93; GTAC with 8 loci: 10.85; scATAC alone: 10.99; Chen plate-based: 11.95; 10X Chromium: 12.33.
(E) Aggregated fragment size distribution across our conditions.
(F) Correlation matrix of Pearson's coefficients comparing GTAC conditions and scATAC alone with published K562 (Chen K562 replicate 1 and Chen K562 replicate 2) and human skin fibroblast scATAC-seq datasets.
(G) Aggregated reads from genome positions chr19:35620000–35790000, across our conditions and plate-based K562 data. Positions of labeled genes are shown.
(H) Percentage of K562 cells passing scATAC QC across our conditions and for the published K562 data (STAR Methods).
(I) Percentage of K562 cells passing both scATAC QC and achieving successful detection of all amplified genotyping loci (STAR Methods).
(J) Percentage of K562 cells passing both scATAC QC and achieving successful detection of all four heterozygous mutations (STAR Methods).
In (K)–(N), 95 cells from each of the three indicated cell lines were subject to GTAC with amplification of 3 genomic loci, *ASXL1*, *RUNX1*, and *NRAS*.
(K) UMAP based on scATAC data from 285 profiled single cells.
(L–N) The UMAP from (K) overlaid with genotyping data from three genomic loci. Legend on the left indicates whether a mutation was detected (red), whether only the WT sequence was detected (blue), or whether there were insufficient reads across the locus (gray). In each panel, the percentage of cells that are false negative (i.e., a mutant cell is called WT) and of cells displaying insufficient reads (detection failure rate) is shown. Overall, 97% of cells were assigned the correct genotype.

CellPress

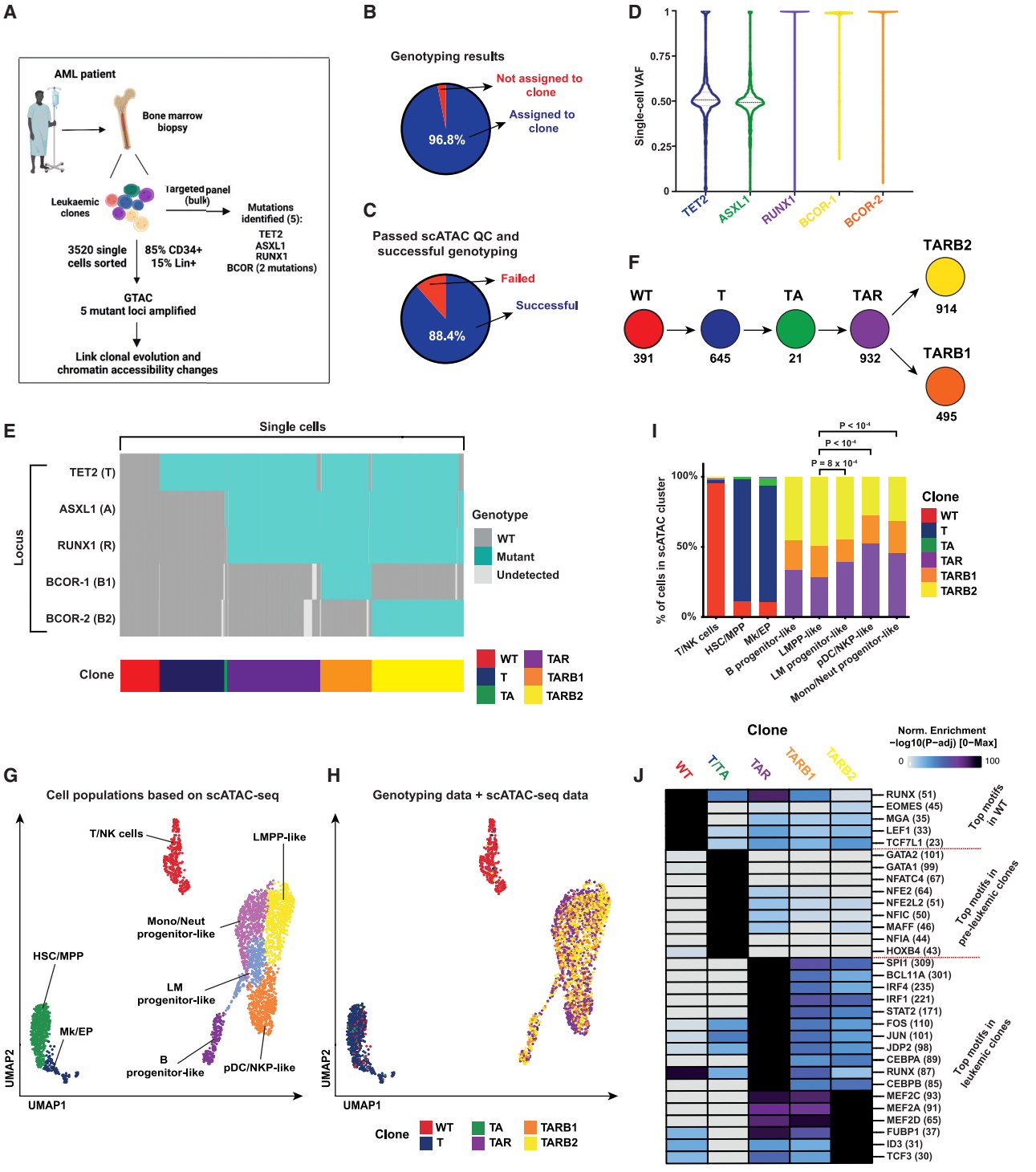


Figure 4. GTAC analysis of primary human AML maps clonal hierarchy onto accessibility profiles and reveals the expansion of specific clones at distinct stages of differentiation

(A) Application of GTAC to 3,520 FACS-sorted single nuclei (85% $CD34^+$ and 15% $Lineage^+$ cells) from a primary human AML sample. Five loci previously identified as mutant by bulk sequencing were targeted.

(B) Percentage of cells assigned to a clone (dark blue, 3,407 of 3,520 cells) based on successful genotyping (STAR Methods).

(C) Percentage of cells passing scATAC QC and assigned to a clone (dark blue, 3,113 of 3,520 cells; STAR Methods).

(legend continued on next page)

stage into account, indicate that leukemic transformation at the CMP/LMPP stage correlates with acquisition of greater binding by TFs known to support LSC function and proliferation.

Taken together, these results demonstrate that in a primary AML sample, GTAC captures the stages of clonal evolution from WT cells through to preleukemia and into the fully transformed AML leukemic stage, showing how distinct clones correlate with different chromatin accessibility landscapes.

GTAC reveals a differentiation delay of *BCOR* mutant clones within the leukemic hierarchy, which biases cells toward an LSC-like chromatin state

Further selection of two independent *BCOR* mutant clones suggests that the acquisition of *BCOR* mutations conferred additional fitness advantage in this somatic background. BCOR is a key component of the repressive non-canonical PRC1.1 complex. Both *BCOR* mutations are likely LoF (both occur in exon 4; one results in a premature stop codon, while the other is a frameshift mutation), suggesting that these may lead to loss of PRC1.1 function in these clones. Compared with other leukemic stages, we noticed higher gene activity of *BCOR* and other PRC1.1 components within the LMPP-like cluster (Figure S5A), where *BCOR* mutant cells accumulate (Figure 4I), suggesting that PRC1.1 may play a role for either entry into or exit from the LMPP-like compartment. Interestingly, the LMPP-like cluster displayed the highest LSC score[43] among all leukemic stages, as well as higher accessibility of several motifs bound by TFs functionally shown to promote self-renewal and AML maintenance[61,63–67] (Figures 5A and 5B). Furthermore, LMPP-like and Mono/Neut progenitor-like cells exhibited higher accessibility of the *TIM-3* gene that discriminates between LSCs and HSCs,[68] and LMPP-like cells had lower accessibility of *CD38* that is downregulated in LSCs in some patient samples[43] (Figure S5B). Consistently, open chromatin within LMPP-like cells was enriched for HOXA and FOXO target genes compared with other leukemic clusters (Figure S5C). Compared with TAR cells, *BCOR* mutant cells were enriched for genes upregulated in HSCs and LSCs (Figure 5C), consistent with *BCOR* mutant cells being enriched for an LSC-like chromatin state.

To characterize dynamic chromatin changes associated with differentiation of specific leukemic subclones, we computed three putative differentiation trajectories undertaken by leukemic cells, from the LMPP-like compartment along the pDC/NKP pathway, Mono/Neut pathway, or the B cell pathway (Figures 5D and S5D–S5F). We observed accumulation of cells from both *BCOR* mutant clones earlier in pseudotime along all trajectories (Figures 5D and S5G), but the distribution of cells across pseudotime was different between cells from the two *BCOR* mutant clones (Figure 5D). The larger *BCOR* mutant clone (TARB2) showed a more pronounced delay in the Mono/Neut and pDC/NKP trajectories compared with TARB1, which was not the case across the leukemic B cell trajectory (Figure S5G). Concordantly, *BCOR* mutant cells showed delayed accessibility of motifs for lineage-affiliated TFs (Figures 5E and S5H). Notably, the magnitude of CEBPA motif accessibility through Mono/Neut pseudotime was lower in *BCOR* mutant cells, and this was particularly pronounced in the TARB2 clone (Figure 5E). Consistent with aberrant myeloid differentiation, there was delayed accessibility of regulatory elements co-accessible with the *TLR4* gene and of chromatin peaks mapping to the *CEBPD* promoter and gene body in TARB2 compared with TAR cells (Figure 5F). Moreover, TARB2 Mono/Neut progenitor-like cells displayed lower accessibility of peaks linked to genes associated with neutrophilic and monocytic differentiation compared with TAR cells (Figure S5I). Conversely, motifs for ETV6, ERG, PBX3, FLI1, YY1, and NF-Y family members, functionally implicated in stem/early progenitor cell self-renewal and/or AML maintenance,[63,65–67,69–72] remained accessible for longer in *BCOR* mutant cells along the pDC/NKP trajectory (Figure 5G). In contrast, there was delayed accessibility of motifs for AP-1 complex members in *BCOR* mutant cells along the same trajectory (Figure 5H). Interestingly, at the early stages of the Mono/Neut pathway, *BCOR* WT cells more quickly gained AP-1 family motif accessibility compared with *BCOR* mutant cells (Figure S5J). The AP-1 family is involved in cell cycle regulation and myelomonocytic differentiation;[73,74] in addition, JUNB exerts a negative effect on leukemic self-renewal capacity[75] and, when overexpressed, leads to loss of HSCs in mice.[76] Indeed, within LMPP-like cells, motif accessibility of JUNB was

(D) scVAF distributions in cells containing the mutation in the indicated genotyped loci. "BCOR-1" and "BCOR-2" refer to the two distinct *BCOR* mutations. Median scVAFs for each locus are: *TET2*: 0.507; *ASXL1*: 0.493; *RUNX1*: 0.997; *BCOR*-1: 0.987; and *BCOR*-2: 0.996. Median and quartile values are shown for *TET2* and *ASXL1*.

(E) Top: single-cell genotyping results. Each row represents a locus (indicated on the left). Each column is a single cell. Colors indicate whether a cell was called WT (gray), mutant (cyan), or undetected (white) for each locus. Bottom: clonal assignment for each cell (color legend on the right). WT, wild-type; T, *TET2* mutant; TA, *TET2/ASXL1* mutant; TAR, *TET2/ASXL1/RUNX1* mutant; TARB1, *TET2/ASXL1/RUNX1/BCOR-1* mutant; TARB2, *TET2/ASXL1/RUNX1/BCOR-2* mutant. Only cells assigned to a clone are displayed.

(F) Most likely clonal structure based on mutation calls across the 5 loci. Numbers of cells assigned to a clone are indicated. Abbreviations are as in (E).

(G) UMAP based on scATAC data of 3,113 single cells from the AML sample. Cluster differentiation stage is defined by chromatin accessibility profiles. HSC/MPP, hematopoietic stem cell/multipotent progenitor; Mk/EP, megakaryocytic/erythroid progenitor; LMPP-like, lymphoid-primed multipotent progenitor-like; Mono/Neut progenitor-like, monocyte/neutrophil progenitor-like; LM progenitor-like, lympho-myeloid progenitor-like; pDC/NKP-like, plasmacytoid dendritic cell/NK progenitor-like.

(H) The UMAP from (G) overlaid with single-cell genotyping data. Clones are labeled below. Abbreviations are as in (F). Cell numbers: WT: 377; T: 576; TA: 19; TAR: 861; TARB2: 834; TARB1: 446.

(I) Percentage of cells from each clone within scATAC clusters. Fisher's exact test was used to assess the frequency of *BCOR* mutant clones between the LMPP-like and other progenitor clusters (significant p values reported above the graph).

(J) Heatmap displaying enrichment of TF motifs (rows) in marker peaks for each clone. Each column is a clone. Colors indicate the normalized motif enrichment ($-\log_{10}$[adjusted p value]) on a scale of 0–100. The maximum motif $-\log_{10}$(adjusted p value) is indicated in brackets. Marker peaks were selected with FDR < 0.05 and $\log_2$FC > 0.5. Only motifs for transcription factors expressed in the hematopoietic system[97,98] are shown.

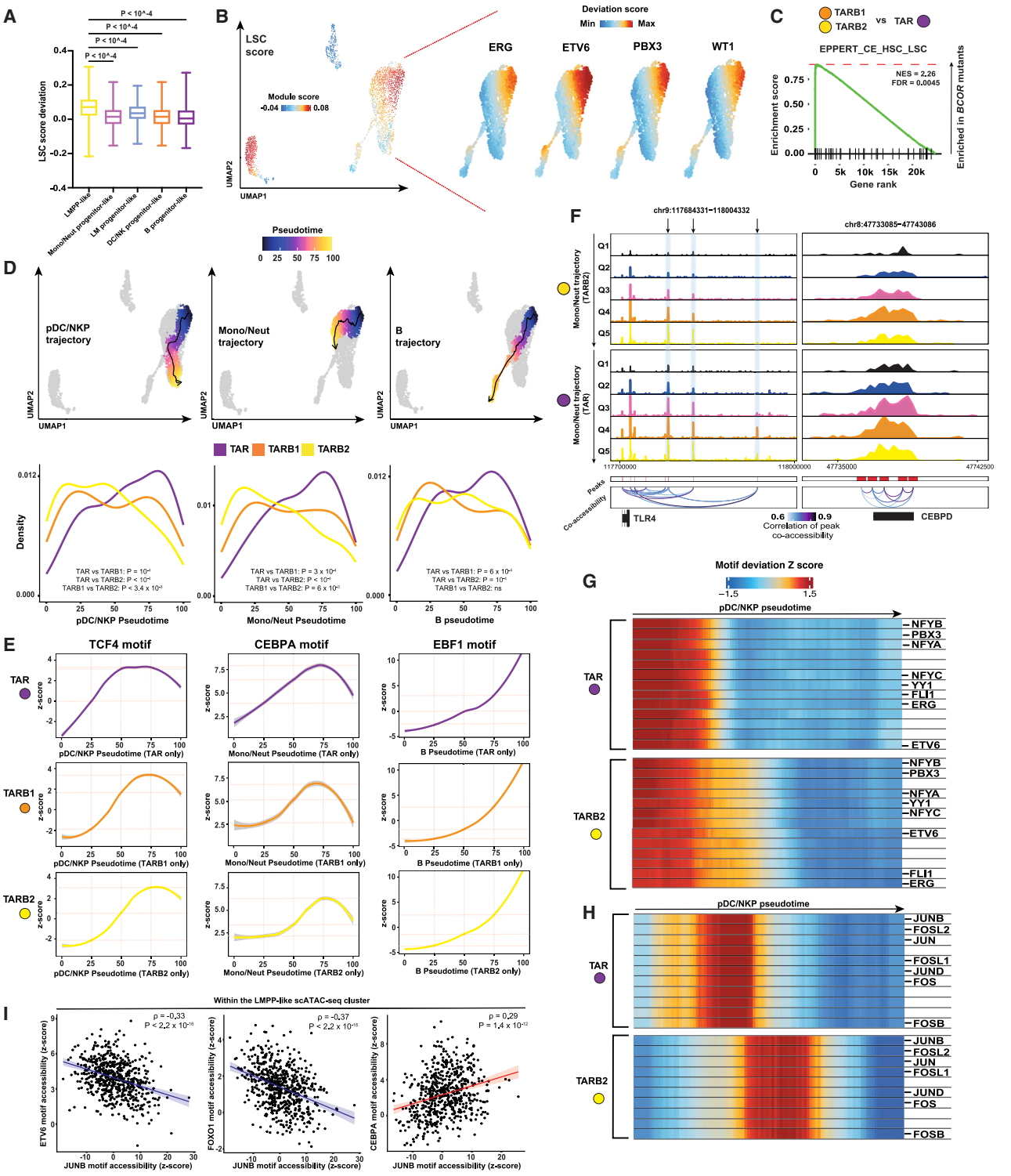


Figure 5. GTAC identifies clone-specific chromatin dynamics throughout the leukemic hierarchy, highlighting a differentiation delay of *BCOR* mutant clones at an LSC-like state

(A) LSC module score deviation from the average of all cells, shown for leukemic populations. Whiskers extend from the minimum to maximum value. Kruskal-Wallis test with Dunnett's multiple comparison was used and p values are reported.

(B) Left: UMAP from Figure 4G overlaid with the LSC module score. Right: leukemic cells overlaid with motif accessibility for ERG, ETV6, PBX3, and WT1.

(C) Gene set enrichment analysis (GSEA) of genes upregulated in HSCs and LSCs compared with committed progenitors.[47] All *BCOR* mutant cells were compared with all TAR cells. Normalized enrichment score (NES) and FDR values are indicated.

(legend continued on next page)

negatively correlated with that of FOXO1 and ETV6, mediators of quiescence and self-renewal, and positively correlated with CEBPA motif accessibility, essential for myeloid differentiation (Figure 5I). Altogether, these data suggest that *BCOR* mutations delay differentiation, biasing leukemic cells toward stem-like chromatin states.

GTAC reveals switches in chromatin accessibility within the LMPP-like differentiation stage upon subclonal evolution of leukemia

Next, we wanted to study chromatin changes related to *BCOR* mutations within the LMPP-like cluster, to provide putative molecular mechanisms for the accumulation of *BCOR* mutant cells at this LSC-like stage. We first compared gene activity between *TET2/ASXL1/RUNX1* mutant *BCOR* WT and *TET2/ASXL1/RUNX1/BCOR* mutant LMPP-like cells (Figures 6A and 6B; Table S2). Notably, *BCOR* mutant LMPP-like cells were enriched in genes upregulated in LSCs and HSCs (Figure 6C). Both *BCOR* mutant clones showed more activity of the *INPP4B*, *SOCS2*, and *NR4A2* genes (Figures 6A and 6B). *INPP4B* and *SOCS2*, putative direct PRC1.1 targets in primary human AML,[77] have been functionally shown to promote HSC and LSC maintenance.[78–80] Increased activity of the *NR4A2* gene, which promotes HSC quiescence,[81] suggests a more quiescent state of *BCOR* mutant LMPP-like cells. Additionally, the larger *BCOR* mutant clone showed higher activity of *PBX3*, *BTG3*, and *JMJD1C* genes, putative PRC1.1 targets related to quiescence and/or promoting leukemogenesis.[47,66,67,82,83] Of note, both INPP4B, an enzyme that dephosphorylates inositol 3,4-biphosphate, and BTG3, which interacts with AKT,[82] are negative regulators of the PI3K/AKT pathway. Conversely, the promoter region and several distal elements linked to *CDK20*, a positive cell cycle regulator, were less accessible in the TARB2 compared with TAR cells (Figure 6D).

We leveraged chromatin information to generate mechanistic hypotheses about aberrant gene regulation associated with *BCOR* LoF. Prior evidence suggests that PRC1.1 binding is strongly biased toward promoter regions.[77,84] We found that peaks more accessible in either TARB2 or TARB1, compared with TAR LMPP-like cells, were enriched for promoters over distal elements, consistent with de-repression of promoters upon loss of PRC1.1 activity (Figure S6A). Consistently, both *BCOR* mutant clones showed increased accessibility of the *INPP4B* promoter region compared with TAR cells (Figures 6E and S6B). Of note, higher *INPP4B* gene activity persisted in *BCOR* mutant cells throughout all leukemic trajectories and was particularly pronounced in the TARB2 clone (Figure 6F). Interestingly, *INPP4B* is among the genes likely to be positively regulated by RUNX1 in mixed-lineage acute leukemia.[41] Consistent with this, *RUNX1* mutant leukemic cells displayed lower *INPP4B* activity compared with preleukemic cells (Figure S6C). It is tempting to speculate that *BCOR* mutations increase fitness advantage in this context by de-repressing the *INPP4B* promoter, compensating for lower *INPP4B* activity in *RUNX1* mutant cells (Figure S6C). In addition, TSS-proximal regions of *NR4A2*, *MLLT10*, *BTG3*, and *SOCS2* genes were also more accessible in *BCOR* mutant cells (Figures 6G and S6D–S6F), and their activity was increased through differentiation in one or both *BCOR* mutant clones (Figures 6H and S6G–S6I).

To better understand how *BCOR* mutations correlated with altered regulatory landscape, we compared TF motif accessibility between *TET2/ASXL1/RUNX1* mutant *BCOR* WT and *TET2/ASXL1/RUNX1/BCOR* mutant LMPP-like cells (Figures 6I and 6J; Table S3). Both *BCOR* mutant clones displayed higher accessibility of HOX family and SMAD motifs, associated with LSC and cancer stem cell function, consistent with previous evidence that *BCOR* mutations are associated with increased expression of *HOX* cluster genes.[85,86] However, there were also unexpected differences between the two *BCOR* mutant clones. TARB1 cells were enriched for ERG, ETV6, and WT1, as well as for MEF and FOXO family members, whereas motifs for NFKB, VENTX, and NANOG were enriched in TARB2 cells. Interestingly, NF-kB signaling was shown to promote HSC self-renewal and survival.[87–89] Accordingly, TF footprinting analysis confirmed more RelA binding in TARB2 LMPP-like cells (Figure S6J). In addition, several other motifs more accessible in one or both *BCOR* mutant clones were bound by TFs previously linked to stemness.[65,70,90,91] Conversely, AP-1, TCF7, IRF4, BCL11 A/B, and CEBPA/D motifs, linked to differentiation and/or cell cycle, were more open in *BCOR* WT LMPP-like cells (Figure 6J).

Finally, to corroborate the unexpected differences between TARB1 and TARB2 clones, we directly compared their gene activity scores across leukemic populations. This revealed increased activity of genes involved in HSC and/or LSC function, including *HOXA9*, *HOXA10*, *KIT*, *BMI1*, *NR4A2*, and *INPP4B*, in the larger TARB2 clone (Figure 6K). Conversely, *CDK20* and *CDC20*, which promote cell cycle, were more accessible in the TARB1 clone. Moreover, binding motifs of IRF4, BCL11 A/B, and SPI1 TFs were also more accessible in the TARB1 clone (Figure S6K). This raises the intriguing possibility that the TARB2 clone had a fitness advantage over the TARB1 clone, occupying a more stem-like chromatin state within the leukemic hierarchy.

(D) Above: UMAP from Figure 4G depicting three differentiation trajectories within leukemia (STAR Methods). Cells are color coded by pseudotime score. Below: for each trajectory, clonal density is plotted as a function of pseudotime. The Kolmogorov-Smirnov test was used to compare distributions. p values for each comparison are indicated.

(E) TCF4, CEBPA, and EBF1 motif accessibility as a function of pseudotime, when each trajectory was split by clone (STAR Methods), indicated on the left.

(F) Normalized scATAC read coverage around the *TLR4* (left) and *CEBPD* (right) genes. Clone-specific Mono/Neut trajectories were split into 5 quantiles from low to high pseudotime (Q1–Q5). Peaks and co-accessibility are plotted below.

(G and H) Heatmaps of motif accessibility across clone-specific pDC/NKP pseudotime (x axis) for a selected subset of most variable motifs (each row represents a motif). In each panel, the upper heatmap relates to TAR and the lower heatmap to TARB2. Motifs bound by TFs discussed in the results section are labeled on the right.

(I) Correlation between JUNB and ETV6 (left), FOXO1 (middle), and CEBPA (right) motif accessibility in LMPP-like cells (dots). Lines represent the linear fit, and shadowed areas represent 95% confidence intervals. Spearman correlation coefficients and p values are reported.

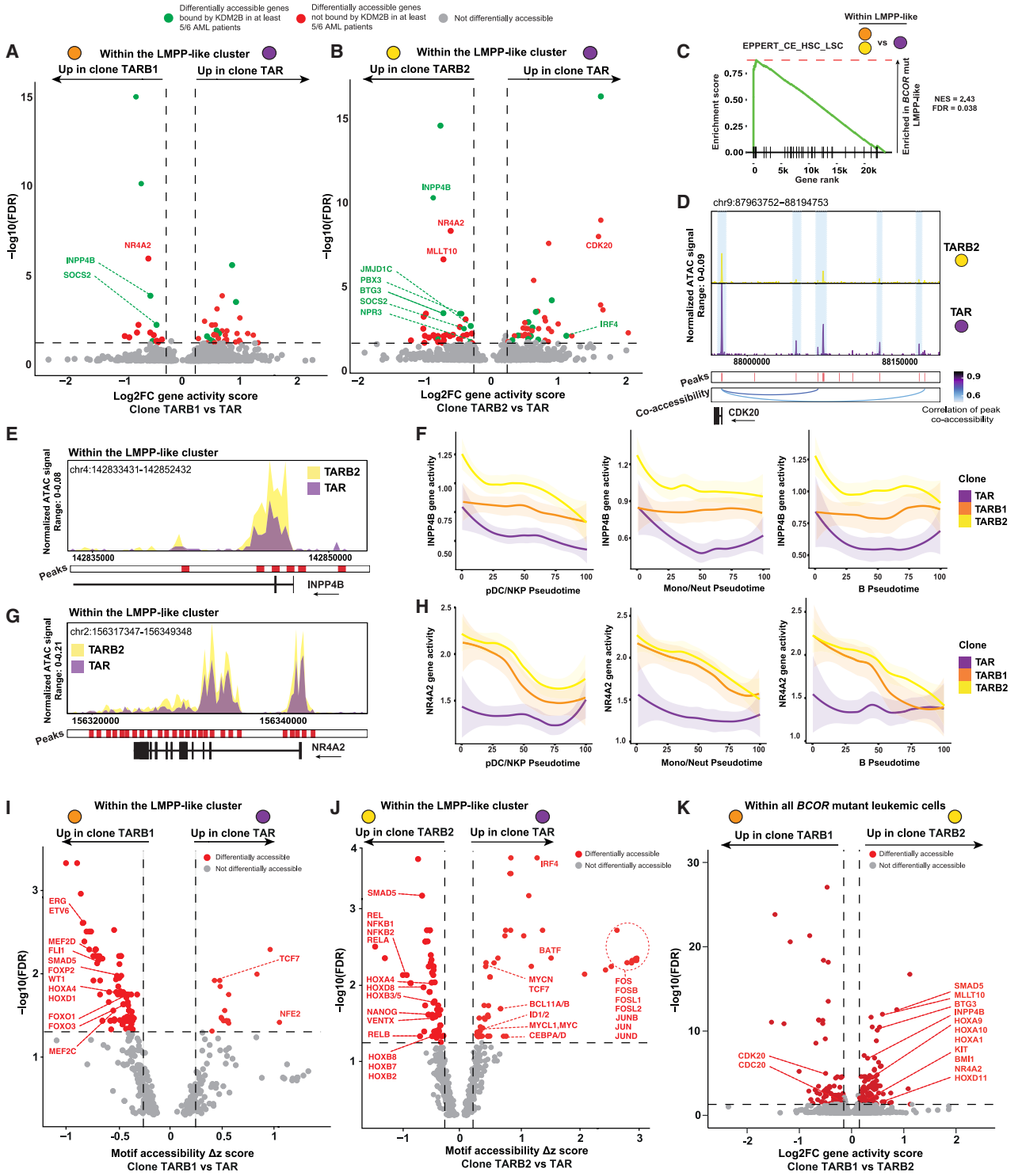


Figure 6. GTAC reveals specific switches in chromatin accessibility within the LMPP-like differentiation stage upon subclonal acquisition of *BCOR* mutations

(A and B) Volcano plots of differential gene activity between TARB1 (n = 130) and TAR (n = 167) clones (A) and TARB2 (n = 293) and TAR clones (B) within the LMPP-like cluster. Differentially accessible genes display FDR < 0.05 and absolute $\log_2$FC > 0.25. Lines indicate $\log_2$FC and FDR thresholds. Differentially accessible putative PRC1.1 targets in AML[77] are labeled in green.

(C) GSEA of genes upregulated in HSCs and LSCs compared with committed progenitors.[47] *BCOR* mutant and WT LMPP-like cells were compared. NES and FDR values are indicated.

(legend continued on next page)

Collectively, the above data suggest that *BCOR* mutations in this somatic context alter the balance between chromatin elements regulating quiescence, cell cycle, and differentiation (Figure S6L). Differential gene and motif openness indicate aberrant reactivation of molecular actors promoting quiescence and maintenance of LSCs at the expense of differentiation. GTAC thus provides specific insights of how mutational cooperativity correlates with chromatin heterogeneity within defined leukemic differentiation stages.

DISCUSSION

We present GTAC, a single-cell multi-modal approach mapping high-quality chromatin accessibility profiles to complex clonal structures in malignant tissues. While our scATAC-seq library construction protocol simplifies and increases cell throughput compared with previous plate-based methods,[33,34] targeted genotyping is easily implementable in any laboratory. Moreover, our analyses adopt broadly utilized computational pipelines, making GTAC a widely accessible method.

Functional assays that rely on either targeted mutagenesis or overexpression in model systems may miss the full breadth of clonal complexity underlying human cancer. Hence, primary samples provide an important benchmark to establish the cellular, epigenetic, and molecular landscape in human pre-malignant and fully transformed clonal populations. While existing approaches demonstrated the correlation of altered genotypes with transcriptomic[17,18] or surface protein heterogeneity,[92] linking somatic mutations to chromatin accessibility leads to several unique advantages. Firstly, chromatin is a more accurate indicator of cell identity compared with RNA,[10] allowing for more granular analysis of cell populations. Secondly, chromatin accessibility enables study of dynamic alterations in TF networks downstream of mutations. Crucially, GTAC reveals proximal or distal regulatory elements with clone-specific activity, enabling mechanistic hypotheses to be generated about the impact of mutations on gene regulation.

We demonstrate the utility of GTAC in human AML, where linking epigenomic profiles to clonal progression identified a *RUNX1* mutation as a key driver event associated with leukemic transformation. Furthermore, more subtle chromatin changes downstream of *BCOR* mutations suggest a de-regulated epigenetic balance between self-renewal and differentiation in *BCOR* mutant progenitors, which gradually expands the pool of LSC-like cells in this patient. Finally, the unexpected difference between two *BCOR* mutant clones suggests that even cells with apparently similar somatic genotypes can display notable chromatin heterogeneity, resulting either from additional unidentified genetic/epigenetic driver(s), from differences in the cells of origin, or due to distinct functional properties of the two mutations. Crucially, GTAC enabled comparison of different clones within specific scATAC-defined populations, a level of granularity not achievable by conventional scATAC approaches. Going forward, GTAC could also be applied to longitudinal patient samples to elucidate clone-specific changes in chromatin profiles throughout treatment, providing insights into epigenetic mechanisms of therapy resistance.

GTAC has important advantages compared with droplet-based technologies.[93] First, high genotyping efficiency and an average ADO of only ~5% per locus make it possible to confidently assign the most likely phylogenetic tree across a range of putative clonal hierarchies, making GTAC uniquely suitable for the study of somatically complex disease. Low ADO rates are crucial to minimize error in downstream analyses, as clone-specific differences in chromatin accessibility would progressively be diluted with increasing ADO rates. Second, locus capture does not depend on chromatin accessibility proximal to the amplified region, allowing for uniform genotyping across different cell states. Third, at parity of Tn5 concentration and cell type, GTAC yields substantially more complex scATAC libraries compared with other technologies,[23–25,93,94] providing greater granularity, which is important when studying subtle differences arising from intra-tumoral heterogeneity. This is currently achieved with a cost per cell of approximately $3 (excluding sequencing). Finally, we speculate that our scATAC library complexity may be further improved by integrating the recently published s3-ATAC strategy.[95]

In conclusion, we provide a robust and highly sensitive single-cell tool capable of resolving chromatin differences between somatic clones within the same tissue. Given that scATAC performs well on a vast spectrum of human cancers,[96] GTAC will be widely applicable across the field of somatic mosaicism and tumor evolution.

LIMITATIONS OF THE STUDY

The main limitation of GTAC is lower throughput compared with droplet-based strategies. In terms of cell numbers, one tagmentation reaction on 50,000 cells usually yields up to ten 384-well plates. This number can be increased depending on starting cell numbers and populations of interest. Timewise, the method requires around 5 days to process ten 384-well plates, but the

(D) Normalized scATAC read coverage around the *CDK20* gene, comparing TARB2 and TAR LMPP-like cells. Peaks and co-accessibility are plotted below. The arrow indicates gene orientation.
(E) Overlaid normalized scATAC read coverage around the *INPP4B* promoter region, comparing TAR and TARB2 LMPP-like cells. Peaks and gene annotation are indicated below. The arrow indicates gene orientation.
(F) *INPP4B* gene activity as a function of pseudotime for the three trajectories. Clones are color coded (legend on the right). The LOESS method was used.
(G) Same as (E) but around the *NR4A2* gene.
(H) Same as (F) but relative to *NR4A2* gene activity.
(I and J) Volcano plots of differential TF motif accessibility between TARB1 and TAR clones (I) and TARB2 and TAR clones (J) within LMPP-like cells. Differentially accessible motifs display FDR < 0.05 and absolute $\Delta z > 0.25$ and are labeled in red. Lines indicate Δz and FDR thresholds.
(K) Volcano plot of differential gene activity between TARB1 (n = 446 cells) and TARB2 (n = 834 cells) clones across all leukemic cells. Differentially accessible genes display FDR < 0.05 and absolute $\log_2$FC > 0.20 and are labeled in red. Lines indicate $\log_2$FC and FDR thresholds.

time required progressively increases with the number of genotyped loci. Altogether, data from 10,000–12,000 cells can be generated in a few weeks. Lower throughput can be limiting when very small clones are studied: in such cases, GTAC is likely underpowered and may miss the full breadth of clone-specific chromatin accessibility. It should also be considered that GTAC relies on the analysis of mutations previously detected in the patient sample by either targeted panel or whole-exome sequencing. Finally, for genetically complex tumors, in which hundreds of mutations are present within the same sample, it would be necessary to limit GTAC analysis to a selected subset of mutations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Patient samples
- METHOD DETAILS
  - Targeted DNA sequencing
  - Primary bone marrow cell sorting
  - GTAC cell lysis and tagmentation
  - GTAC lysis plate preparation, single-nucleus sorting and preamplification
  - GTAC single-nucleus genotyping library preparation
  - GTAC sequencing
  - GTAC primer design and validation
  - Single-cell genotyping data pre-processing and mutation calling
  - Inference of clonal hierarchies
  - Allelic dropout (ADO) estimation
  - Coverage calculation from scATAC data
  - scATAC data pre-processing
  - Correlation of peak calls
  - Merging scATAC and genotyping data
  - Doublet exclusion
  - Dimensionality reduction and clustering
  - Gene activity and module score
  - Pseudo-bulk replicate generation, peak calling, and marker identification
  - Differential accessibility analysis and motif enrichment
  - Co-accessibility
  - Projection analysis
  - Trajectory analysis
  - Motif correlation analysis
  - Footprinting analysis
  - Enrichment of genomic elements in differentially accessible peaks
  - Identification of putative PRC1.1 direct targets in AML from public data
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2023.04.012.

ACKNOWLEDGMENTS

P.V. acknowledges funding from the Medical Research Council Molecular Haematology Unit Programme Grant (MC_UU_00029/8), Blood Cancer UK Programme Continuity Grant 13008, NIHR Senior Fellowship, and the Oxford BRC Haematology Theme. S.T. is supported by a Scatcherd European Scholarship in partnership with The Medical Research Council/Radcliffe Department of Medicine and The Clarendon Fund. N.A.J. is supported by a Medical Research Council and Leukaemia UK Clinical Research Training Fellowship (MR/R002258/1). A.G. is supported by DPhil studentship funding from the Lady Tata Memorial Trust. S.G.R., E.R.G., and J.R.H. acknowledge funding from the Medical Research Council (MC_UU_00029/3) and a Wellcome Trust Strategic Award (106130/Z/14/Z). E.R.G. is supported by "The Ministry of National Education Selection and Placement of Candidates Sent Abroad for Postgraduate Education (YLSY) scholarship," Republic of Türkiye, Ministry of National Education. L.D.H. was supported by the National Institutes of Health (USA) (R24DK106766). G.M. is supported by a Cancer Research UK DPhil studentship (C2195/A31310). J.O.J.D. is supported by a Medical Research Council Clinician Scientist Fellowship (MR/R008108) and the Medical Research Council Molecular Haematology Unit Programme Grant (MC_UU_00029/4) and the Lister Institute Fellowship. The authors thank Prof. Ketan Patel, Prof. Thomas Milne, Dr. Sally-Ann Clark, Dr. Bilyana Stoilova, Dr. David Cruz Hernandez, and Mr. Felix Radtke for insightful comments and discussions. The authors also acknowledge the MRC WIMM Flow Cytometry Facility and Single Cell Facility. Some of the figures in this manuscript were created using BioRender.

AUTHOR CONTRIBUTIONS

Conceptualization, S.T., N.A.J., A.G., and P.V.; methodology, S.T., N.A.J., and A.G.; investigation, S.T., N.A.J., A.G., M.M., B.U., M.A.S., G.M., K.C., and P.S.; formal analysis, S.T., N.A.J., S.G.R., E.R.G., and L.D.H.; funding acquisition, P.V.; supervision, J.O.J.D., J.R.H., and P.V.; writing – original draft, S.T., N.A.J., and P.V.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

J.O.J.D. and J.R.H. are co-founders of Nucleome Therapeutics and provide consultancy to the company.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: October 4, 2022
Revised: February 23, 2023
Accepted: April 12, 2023
Published: May 4, 2023

REFERENCES

1. Forsberg, L.A., Gisselsson, D., and Dumanski, J.P. (2017). Mosaicism in health and disease - clones picking up speed. Nat. Rev. Genet. *18*, 128–142. https://doi.org/10.1038/nrg.2016.145.
2. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. *371*, 2488–2498. https://doi.org/10.1056/NEJMoa1408617.
3. Bowman, R.L., Busque, L., and Levine, R.L. (2018). Clonal hematopoiesis and evolution to hematopoietic malignancies. Cell Stem Cell *22*, 157–170. https://doi.org/10.1016/j.stem.2018.01.011.

4. Sperling, A.S., Gibson, C.J., and Ebert, B.L. (2017). The genetics of myelodysplastic syndrome: From clonal haematopoiesis to secondary leukaemia. Nat. Rev. Cancer *17*, 5–19. https://doi.org/10.1038/nrc.2016.112.

5. Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. Science *349*, 1483–1489. https://doi.org/10.1126/science.aab4082.

6. McGranahan, N., and Swanton, C. (2017). Clonal heterogeneity and tumor evolution: Past, present, and the future. Cell *168*, 613–628. https://doi.org/10.1016/j.cell.2017.01.018.

7. Woyach, J.A., Furman, R.R., Liu, T.M., Ozer, H.G., Zapatka, M., Ruppert, A.S., Xue, L., Li, D.H.H., Steggerda, S.M., Versele, M., et al. (2014). Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. N. Engl. J. Med. *370*, 2286–2294. https://doi.org/10.1056/NEJMoa1400029.

8. Quek, L., David, M.D., Kennedy, A., Metzner, M., Amatangelo, M., Shih, A., Stoilova, B., Quivoron, C., Heiblig, M., Willekens, C., et al. (2018). Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib. Nat. Med. *24*, 1167–1177. https://doi.org/10.1038/s41591-018-0115-6.

9. Morita, K., Wang, F., Jahn, K., Hu, T., Tanaka, T., Sasaki, Y., Kuipers, J., Loghavi, S., Wang, S.A., Yan, Y., et al. (2020). Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. Nat. Commun. *11*, 5327. https://doi.org/10.1038/s41467-020-19119-8.

10. Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203. https://doi.org/10.1038/ng.3646.

11. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220. https://doi.org/10.1038/s41576-018-0089-8.

12. Ushijima, T., Clark, S.J., and Tan, P. (2021). Mapping genomic and epigenomic evolution in cancer ecosystems. Science *373*, 1474–1479. https://doi.org/10.1126/science.abh1645.

13. Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat. Genet. *52*, 378–387. https://doi.org/10.1038/s41588-020-0595-4.

14. Nam, A.S., Dusaj, N., Izzo, F., Murali, R., Myers, R.M., Mouhieddine, T., Sotelo, J., Benbarche, S., Waarts, M., Gaiti, F., et al. (2022). Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.14.476225.

15. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93. https://doi.org/10.1038/s41586-020-1969-6.

16. Cancer Genome Atlas Research Network, Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A.G., Hoadley, K., Triche, T.J., Jr., Laird, P.W., et al. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074. https://doi.org/10.1056/NEJMoa1301689.

17. Rodriguez-Meira, A., Buck, G., Clark, S.A., Povinelli, B.J., Alcolea, V., Louka, E., McGowan, S., Hamblin, A., Sousos, N., Barkas, N., et al. (2019). Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. Mol. Cell *73*, 1292–1305.e8. https://doi.org/10.1016/j.molcel.2019.01.009.

18. Nam, A.S., Kim, K.T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R.M., Abu-Zeinah, G., Brand, R., Omans, N.D., et al. (2019). Somatic mutations and cell identity linked by genotyping of transcriptomes. Nature *571*, 355–360. https://doi.org/10.1038/s41586-019-1367-0.

19. Velten, L., Story, B.A., Hernández-Malmierca, P., Raffel, S., Leonce, D.R., Milbank, J., Paulsen, M., Demir, A., Szu-Tu, C., Frömel, R., et al. (2021). Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. Nat. Commun. *12*, 1366. https://doi.org/10.1038/s41467-021-21650-1.

20. Rodriguez-Meira, A., Norfo, R., Wen, W.X., Chédeville, A.L., Rahman, H., O'Sullivan, J., Wang, G., Louka, E., Kretzschmar, W.W., Paterson, A., et al. (2022). Deciphering TP53 mutant cancer evolution with single-cell multi-omics. Preprint at bioRxiv. https://doi.org/10.1101/2022.03.28.485984.

21. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

22. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962. https://doi.org/10.1038/nmeth.4396.

23. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490. https://doi.org/10.1038/nature14590.

24. Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol. *37*, 925–936. https://doi.org/10.1038/s41587-019-0206-z.

25. Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S., Pokholok, D., Aryee, M.J., Steemers, F.J., Lebofsky, R., and Buenrostro, J.D. (2019). Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat. Biotechnol. *37*, 916–924. https://doi.org/10.1038/s41587-019-0147-6.

26. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell *183*, 1103–1116.e20. https://doi.org/10.1016/j.cell.2020.09.056.

27. Xu, W., Yang, W., Zhang, Y., Chen, Y., Hong, N., Zhang, Q., Wang, X., Hu, Y., Song, K., Jin, W., and Chen, X. (2022). ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.16.476488.

28. Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science *361*, 1380–1385. https://doi.org/10.1126/science.aau0730.

29. Cheng, H., Pui, H.-p., Lentini, A., Kolbeinsdóttir, S., Andrews, N., Pei, Y., Reinius, B., Deng, Q., and Enge, M. (2021). Smart3-ATAC: A highly sensitive method for joint accessibility and full-length transcriptome analysis in single cells. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.02.470912.

30. Mimitou, E.P., Lareau, C.A., Chen, K.Y., Zorzetto-Fernandes, A.L., Hao, Y., Takeshima, Y., Luo, W., Huang, T.S., Yeung, B.Z., Papalexi, E., et al. (2021). Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat. Biotechnol. *39*, 1246–1258. https://doi.org/10.1038/s41587-021-00927-2.

31. Chen, A.F., Parks, B., Kathiria, A.S., Ober-Reynolds, B., Goronzy, J.J., and Greenleaf, W.J. (2022). NEAT-seq: Simultaneous profiling of intranuclear proteins, chromatin accessibility and gene expression in single cells. Nat. Methods *19*, 547–553. https://doi.org/10.1038/s41592-022-01461-y.

32. Lareau, C.A., Ludwig, L.S., Muus, C., Gohil, S.H., Zhao, T., Chiang, Z., Pelka, K., Verboon, J.M., Luo, W., Christian, E., et al. (2021). Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. Nat. Biotechnol. *39*, 451–461. https://doi.org/10.1038/s41587-020-0645-6.

33. Chen, X., Miragaia, R.J., Natarajan, K.N., and Teichmann, S.A. (2018). A rapid and robust method for single cell chromatin accessibility profiling. Nat. Commun. *9*, 5345. https://doi.org/10.1038/s41467-018-07771-0.

34. Xu, W., Wen, Y., Liang, Y., Xu, Q., Wang, X., Jin, W., and Chen, X. (2021). A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility. Nat. Protoc. *16*, 4084–4107. https://doi.org/10.1038/s41596-021-00583-5.

35. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947. https://doi.org/10.1093/nar/gky1015.

36. Pierce, S.E., Granja, J.M., and Greenleaf, W.J. (2021). High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Nat. Commun. *12*, 2969. https://doi.org/10.1038/s41467-021-23213-w.

37. Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: Advances through genomics. Nat. Rev. Genet. *10*, 161–172. https://doi.org/10.1038/nrg2522.

38. Bravo, J., Li, Z., Speck, N.A., and Warren, A.J. (2001). The leukemia-associated AML1 (Runx1)–CBFβ complex functions as a DNA-induced molecular clamp. Nat. Struct. Biol. *8*, 371–378. https://doi.org/10.1038/86264.

39. Ullah, H., Zhang, B., Sharma, N.K., McCrea, P.D., and Srivastava, Y. (2022). In-silico probing of AML related RUNX1 cancer-associated missense mutations: Predicted relationships to DNA binding and drug interactions. Front. Mol. Biosci. *9*, 981020. https://doi.org/10.3389/fmolb.2022.981020.

40. Song, W.-J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. Nat. Genet. *23*, 166–175. https://doi.org/10.1038/13793.

41. Granja, J.M., Klemm, S., McGinnis, L.M., Kathiria, A.S., Mezger, A., Corces, M.R., Parks, B., Gars, E., Liedtke, M., Zheng, G.X.Y., et al. (2019). Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat. Biotechnol. *37*, 1458–1465. https://doi.org/10.1038/s41587-019-0332-7.

42. Ficara, F., Murphy, M.J., Lin, M., and Cleary, M.L. (2008). Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. Cell Stem Cell *2*, 484–496. https://doi.org/10.1016/j.stem.2008.03.004.

43. Gentles, A.J., Plevritis, S.K., Majeti, R., and Alizadeh, A.A. (2010). Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. JAMA *304*, 2706–2715. https://doi.org/10.1001/jama.2010.1862.

44. Corces-Zimmerman, M.R., Hong, W.-J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc. Natl. Acad. Sci. USA *111*, 2548–2553. https://doi.org/10.1073/pnas.1324297111.

45. Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Sci. Transl. Med. *4*, 149ra118. https://doi.org/10.1126/scitranslmed.3004315.

46. Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R., et al. (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell *19*, 138–152. https://doi.org/10.1016/j.ccr.2010.12.012.

47. Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J., et al. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nat. Med. *17*, 1086–1093. https://doi.org/10.1038/nm.2415.

48. Shlush, L.I., Mitchell, A., Heisler, L., Abelson, S., Ng, S.W.K., Trotman-Grant, A., Medeiros, J.J.F., Rao-Bhatia, A., Jaciw-Zurakowsky, I., Marke, R., et al. (2017). Tracing the origins of relapse in acute myeloid leukaemia to stem cells. Nature *547*, 104–108. https://doi.org/10.1038/nature22993.

49. Mendler, J.H., Maharry, K., Radmacher, M.D., Mrózek, K., Becker, H., Metzeler, K.H., Schwind, S., Whitman, S.P., Khalife, J., Kohlschmidt, J., et al. (2012). RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. J. Clin. Oncol. *30*, 3109–3118. https://doi.org/10.1200/JCO.2011.40.6652.

50. Greif, P.A., Konstandin, N.P., Metzeler, K.H., Herold, T., Pasalic, Z., Ksienzyk, B., Dufour, A., Schneider, F., Schneider, S., Kakadia, P.M., et al. (2012). RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes. Haematologica *97*, 1909–1915. https://doi.org/10.3324/haematol.2012.064667.

51. Heuser, M., Beutel, G., Krauter, J., Döhner, K., von Neuhoff, N., Schlegelberger, B., and Ganser, A. (2006). High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. Blood *108*, 3898–3905. https://doi.org/10.1182/blood-2006-04-014845.

52. Intlekofer, A.M., Takemoto, N., Wherry, E.J., Longworth, S.A., Northrup, J.T., Palanivel, V.R., Mullen, A.C., Gasink, C.R., Kaech, S.M., Miller, J.D., et al. (2005). Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. Nat. Immunol. *6*, 1236–1244. https://doi.org/10.1038/ni1268.

53. Paley, M.A., Kroy, D.C., Odorizzi, P.M., Johnnidis, J.B., Dolfi, D.V., Barnett, B.E., Bikoff, E.K., Robertson, E.J., Lauer, G.M., Reiner, S.L., and Wherry, E.J. (2012). Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. Science *338*, 1220–1225. https://doi.org/10.1126/science.1229620.

54. Okamura, R.M., Sigvardsson, M., Galceran, J., Verbeek, S., Clevers, H., and Grosschedl, R. (1998). Redundant regulation of T cell differentiation and TCRalpha gene expression by the transcription factors LEF-1 and TCF-1. Immunity *8*, 11–20. https://doi.org/10.1016/s1074-7613(00)80454-9.

55. Wong, W.F., Kohu, K., Chiba, T., Sato, T., and Satake, M. (2011). Interplay of transcription factors in T-cell differentiation and function: The role of Runx. Immunology *132*, 157–164. https://doi.org/10.1111/j.1365-2567.2010.03381.x.

56. Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: Balancing differentiation with transformation. Nat. Rev. Immunol. *7*, 105–117. https://doi.org/10.1038/nri2024.

57. Yu, Y., Wang, J., Khaled, W., Burke, S., Li, P., Chen, X., Yang, W., Jenkins, N.A., Copeland, N.G., Zhang, S., and Liu, P. (2012). Bcl11a is essential for lymphoid development and negatively regulates p53. J. Exp. Med. *209*, 2467–2483. https://doi.org/10.1084/jem.20121846.

58. Brown, F.C., Still, E., Koche, R.P., Yim, C.Y., Takao, S., Cifani, P., Reed, C., Gunasekera, S., Ficarro, S.B., Romanienko, P., et al. (2018). MEF2C phosphorylation is required for chemotherapy resistance in acute myeloid leukemia. Cancer Discov. *8*, 478–497. https://doi.org/10.1158/2159-8290.CD-17-1271.

59. Schwieger, M., Schüler, A., Forster, M., Engelmann, A., Arnold, M.A., Delwel, R., Valk, P.J., Löhler, J., Slany, R.K., Olson, E.N., and Stocking, C. (2009). Homing and invasiveness of MLL/ENL leukemic cells is regulated by MEF2C. Blood *114*, 2476–2488. https://doi.org/10.1182/blood-2008-05-158196.

60. Zhao, L., Zhang, P., Galbo, P.M., Zhou, X., Aryal, S., Qiu, S., Zhang, H., Zhou, Y., Li, C., Zheng, D., et al. (2021). Transcription factor MEF2D is required for the maintenance of MLL-rearranged acute myeloid leukemia. Blood Adv. *5*, 4727–4740. https://doi.org/10.1182/bloodadvances.2021004469.

61. Rampal, R., and Figueroa, M.E. (2016). Wilms tumor 1 mutations in the pathogenesis of acute myeloid leukemia. Haematologica *101*, 672–679. https://doi.org/10.3324/haematol.2015.141796.

62. Wong, P., Iwasaki, M., Somervaille, T.C.P., So, C.W.E., and Cleary, M.L. (2007). Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. Genes Dev. *21*, 2762–2774. https://doi.org/10.1101/gad.1602107.

63. Hock, H., Meade, E., Medeiros, S., Schindler, J.W., Valk, P.J.M., Fujiwara, Y., and Orkin, S.H. (2004). Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. Genes Dev. *18*, 2336–2341. https://doi.org/10.1101/gad.1239604.

64. Guo, H., Chu, Y., Wang, L., Chen, X., Chen, Y., Cheng, H., Zhang, L., Zhou, Y., Yang, F.-c., Cheng, T., et al. (2017). PBX3 is essential for leukemia stem cell maintenance in MLL-rearranged leukemia. Int. J. Cancer *141*, 324–335. https://doi.org/10.1002/ijc.30739.

65. Knudsen, K.J., Rehn, M., Hasemann, M.S., Rapin, N., Bagger, F.O., Ohlsson, E., Willer, A., Frank, A.-K., Søndergaard, E., Jendholm, J., et al. (2015). ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. Genes Dev. *29*, 1915–1929. https://doi.org/10.1101/gad.268409.115.

66. Li, Z., Chen, P., Su, R., Hu, C., Li, Y., Elkahloun, A.G., Zuo, Z., Gurbuxani, S., Arnovitz, S., Weng, H., et al. (2016). PBX3 and MEIS1 cooperate in hematopoietic cells to drive acute myeloid leukemias characterized by a core transcriptome of the MLL-rearranged disease. Cancer Res. *76*, 619–629. https://doi.org/10.1158/0008-5472.Can-15-1566.

67. Li, Z., Zhang, Z., Li, Y., Arnovitz, S., Chen, P., Huang, H., Jiang, X., Hong, G.-M., Kunjamma, R.B., Ren, H., et al. (2013). PBX3 is an important cofactor of HOXA9 in leukemogenesis. Blood *121*, 1422–1431. https://doi.org/10.1182/blood-2012-07-442004.

68. Jan, M., Chao, M.P., Cha, A.C., Alizadeh, A.A., Gentles, A.J., Weissman, I.L., and Majeti, R. (2011). Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. Proc. Natl. Acad. Sci. USA *108*, 5009–5014. https://doi.org/10.1073/pnas.1100551108.

69. He, X., Xu, Y., Huang, D., Yu, Z., Yu, J., Xie, L., Liu, L., Yu, Y., Chen, C., Wan, J., et al. (2023). P2X1 enhances leukemogenesis through PBX3-BCAT1 pathways. Leukemia *37*, 265–275. https://doi.org/10.1038/s41375-022-01759-y.

70. Kruse, E.A., Loughran, S.J., Baldwin, T.M., Josefsson, E.C., Ellis, S., Watson, D.K., Nurden, P., Metcalf, D., Hilton, D.J., Alexander, W.S., and Kile, B.T. (2009). Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage. Proc. Natl. Acad. Sci. USA *106*, 13814–13819. https://doi.org/10.1073/pnas.0906556106.

71. Lu, Z., Hong, C.C., Kong, G., Assumpção, A.L.F.V., Ong, I.M., Bresnick, E.H., Zhang, J., and Pan, X. (2018). Polycomb group protein yy1 is an essential regulator of hematopoietic stem cell quiescence. Cell Rep. *22*, 1545–1559. https://doi.org/10.1016/j.celrep.2018.01.026.

72. Bungartz, G., Land, H., Scadden, D.T., and Emerson, S.G. (2012). NF-Y is necessary for hematopoietic stem cell proliferation and survival. Blood *119*, 1380–1389. https://doi.org/10.1182/blood-2011-06-359406.

73. Behre, G., Whitmarsh, A.J., Coghlan, M.P., Hoang, T., Carpenter, C.L., Zhang, D.-E., Davis, R.J., and Tenen, D.G. (1999). c-Jun is a JNK-independent coactivator of the PU.1 transcription factor *. J. Biol. Chem. *274*, 4939–4946. https://doi.org/10.1074/jbc.274.8.4939.

74. Shafarenko, M., Amanullah, A., Gregory, B., Liebermann, D.A., and Hoffman, B. (2004). Fos modulates myeloid cell survival and differentiation and partially abrogates the c-Myc block in terminal myeloid differentiation. Blood *103*, 4259–4267. https://doi.org/10.1182/blood-2002-09-2704.

75. Steidl, U., Rosenbauer, F., Verhaak, R.G.W., Gu, X., Ebralidze, A., Otu, H.H., Klippel, S., Steidl, C., Bruns, I., Costa, D.B., et al. (2006). Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. Nat. Genet. *38*, 1269–1277. https://doi.org/10.1038/ng1898.

76. Passegué, E., Wagner, E.F., and Weissman, I.L. (2004). JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. Cell *119*, 431–443. https://doi.org/10.1016/j.cell.2004.10.010.

77. van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W.J., Yuan, H., et al. (2016). Non-canonical PRC1.1 targets active genes independent of H3K27me3 and is essential for leukemogenesis. Cell Rep. *14*, 332–346. https://doi.org/10.1016/j.celrep.2015.12.034.

78. Woolley, J.F., Chen, K., Genc, G.E., Lee, D.K.C., Dzneladze, I., He, R., Gabra, M.M., Saffi, G.T., Son, M.H., Schoof, E.M., et al. (2021). INPP4B promotes leukemia by restricting leukemic stem cell differentiation through regulation of lysosomal functions. Preprint at bioRxiv. https://doi.org/10.1101/2021.03.25.437029.

79. Nguyen, C.H., Glüxam, T., Schlerka, A., Bauer, K., Grandits, A.M., Hackl, H., Dovey, O., Zöchbauer-Müller, S., Cooper, J.L., Vassiliou, G.S., et al. (2019). SOCS2 is part of a highly prognostic 4-gene signature in AML and promotes disease aggressiveness. Sci. Rep. *9*, 9139. https://doi.org/10.1038/s41598-019-45579-0.

80. Vitali, C., Bassani, C., Chiodoni, C., Fellini, E., Guarnotta, C., Miotti, S., Sangaletti, S., Fuligni, F., De Cecco, L., Piccaluga, P.P., et al. (2015). SOCS2 controls proliferation and stemness of hematopoietic cells under stress conditions and its deregulation marks unfavorable acute leukemias. Cancer Res. *75*, 2387–2399. https://doi.org/10.1158/0008-5472.Can-14-3625.

81. Sirin, O., Lukov, G.L., Mao, R., Conneely, O.M., and Goodell, M.A. (2010). The orphan nuclear receptor Nurr1 restricts the proliferation of haematopoietic stem cells. Nat. Cell Biol. *12*, 1213–1219. https://doi.org/10.1038/ncb2125.

82. Cheng, Y.C., Chen, P.H., Chiang, H.Y., Suen, C.S., Hwang, M.J., Lin, T.Y., Yang, H.C., Lin, W.C., Lai, P.L., and Shieh, S.Y. (2015). Candidate tumor suppressor B-cell translocation gene 3 impedes neoplastic progression by suppression of AKT. Cell Death Dis. *6*, e1584. https://doi.org/10.1038/cddis.2014.550.

83. Chen, M., Zhu, N., Liu, X., Laurent, B., Tang, Z., Eng, R., Shi, Y., Armstrong, S.A., and Roeder, R.G. (2015). JMJD1C is required for the survival of acute myeloid leukemia by functioning as a coactivator for key transcription factors. Genes Dev. *29*, 2123–2139. https://doi.org/10.1101/gad.267278.115.

84. Wang, Z., Gearhart, M.D., Lee, Y.-W., Kumar, I., Ramazanov, B., Zhang, Y., Hernandez, C., Lu, A.Y., Neuenkirchen, N., Deng, J., et al. (2018). A non-canonical BCOR-PRC1.1 complex represses differentiation programs in human ESCs. Cell Stem Cell *22*, 235–251.e9. https://doi.org/10.1016/j.stem.2017.12.002.

85. Cao, Q., Gearhart, M.D., Gery, S., Shojaee, S., Yang, H., Sun, H., Lin, D.c., Bai, J.w., Mead, M., Zhao, Z., et al. (2016). BCOR regulates myeloid cell proliferation and differentiation. Leukemia *30*, 1155–1165. https://doi.org/10.1038/leu.2016.2.

86. Kelly, M.J., So, J., Rogers, A.J., Gregory, G., Li, J., Zethoven, M., Gearhart, M.D., Bardwell, V.J., Johnstone, R.W., Vervoort, S.J., and Kats, L.M. (2019). Bcor loss perturbs myeloid differentiation and promotes leukaemogenesis. Nat. Commun. *10*, 1347. https://doi.org/10.1038/s41467-019-09250-6.

87. Stein, S.J., and Baldwin, A.S. (2013). Deletion of the NF-κB subunit p65/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function. Blood *121*, 5015–5024. https://doi.org/10.1182/blood-2013-02-486142.

88. Zhao, C., Xiu, Y., Ashton, J., Xing, L., Morita, Y., Jordan, C.T., and Boyce, B.F. (2012). Noncanonical NF-κB signaling regulates hematopoietic stem cell self-renewal and microenvironment interactions. Stem Cell. *30*, 709–718. https://doi.org/10.1002/stem.1050.

89. Yamashita, M., and Passegué, E. (2019). TNF-α coordinates hematopoietic stem cell survival and myeloid regeneration. Cell Stem Cell *25*, 357–372.e7. https://doi.org/10.1016/j.stem.2019.05.019.

90. Tothova, Z., Kollipara, R., Huntly, B.J., Lee, B.H., Castrillon, D.H., Cullen, D.E., McDowell, E.P., Lazo-Kallanian, S., Williams, I.R., Sears, C., et al. (2007). FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. Cell *128*, 325–339. https://doi.org/10.1016/j.cell.2007.01.003.

91. Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia *27*, 1000–1008. https://doi.org/10.1038/leu.2012.356.

92. Miles, L.A., Bowman, R.L., Merlinsky, T.R., Csete, I.S., Ooi, A.T., Durruthy-Durruthy, R., Bowman, M., Famulare, C., Patel, M.A., Mendez, P., et al. (2020). Single-cell mutation analysis of clonal evolution in myeloid malignancies. Nature *587*, 477–482. https://doi.org/10.1038/s41586-020-2864-x.

93. Myers, R.M., Izzo, F., Kottapalli, S., Prieto, T., Dunbar, A., Bowman, R.L., Mimitou, E.P., Stahl, M., El Ghaity-Beckley, S., Arandela, J., et al. (2022). Integrated single-cell genotyping and chromatin accessibility charts JAK2 V617F human hematopoietic differentiation. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.11.491515.

94. Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell *173*, 1535–1548.e16. https://doi.org/10.1016/j.cell.2018.03.074.

95. Mulqueen, R.M., Pokholok, D., O'Connell, B.L., Thornton, C.A., Zhang, F., O'Roak, B.J., Link, J., Yardımcı, G.G., Sears, R.C., Steemers, F.J., and Adey, A.C. (2021). High-content single-cell combinatorial indexing. Nat. Biotechnol. *39*, 1574–1580. https://doi.org/10.1038/s41587-021-00962-z.

96. Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898. https://doi.org/10.1126/science.aav1898.

97. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Sci. Adv. *7*, eabh2169. https://doi.org/10.1126/sciadv.abh2169.

98. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

99. Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic classification and prognosis in acute myeloid leukemia. N. Engl. J. Med. *374*, 2209–2221. https://doi.org/10.1056/NEJMoa1516192.

100. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. *371*, 2477–2487. https://doi.org/10.1056/NEJMoa1409405.

101. Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. *20*, 1472–1478. https://doi.org/10.1038/nm.3733.

102. Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S.H., Kiemeney, L.A.L.M., Veltman, J.A., Gilissen, C., and Hoischen, A. (2017). Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. Am. J. Hum. Genet. *101*, 50–64. https://doi.org/10.1016/j.ajhg.2017.05.013.

103. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

104. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson, J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. *44*, e108. https://doi.org/10.1093/nar/gkw227.

105. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.

106. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. *10*, 1556–1566. https://doi.org/10.1038/nprot.2015.105.

107. Rodriguez-Meira, A., O'Sullivan, J., Rahman, H., and Mead, A.J. (2020). TARGET-Seq: A protocol for high-sensitivity single-cell mutational analysis and parallel RNA sequencing. STAR Protoc. *1*, 100125. https://doi.org/10.1016/j.xpro.2020.100125.

108. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinf. *13*, 134. https://doi.org/10.1186/1471-2105-13-134.

109. Blätte, T.J., Schmalbrock, L.K., Skambraks, S., Lux, S., Cocciardi, S., Dolnik, A., Döhner, H., Döhner, K., and Bullinger, L. (2019). getITD for FLT3-ITD-based MRD monitoring in AML. Leukemia *33*, 2535–2539. https://doi.org/10.1038/s41375-019-0483-z.

110. Jahn, K., Kuipers, J., and Beerenwinkel, N. (2016). Tree inference for single-cell data. Genome Biol. *17*, 86. https://doi.org/10.1186/s13059-016-0936-x.

111. Hentges, L.D., Sergeant, M.J., Cole, C.B., Downes, D.J., Hughes, J.R., and Taylor, S. (2022). LanceOtron: A deep learning peak caller for genome sequencing experiments. Bioinformatics *38*, 4255–4263. https://doi.org/10.1093/bioinformatics/btac525.

112. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.

113. Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature *574*, 365–371. https://doi.org/10.1038/s41586-019-1652-y.

114. Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., et al. (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science *345*, 1251033. https://doi.org/10.1126/science.1251033.

115. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., et al. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nat. Immunol. *19*, 85–97. https://doi.org/10.1038/s41590-017-0001-2.

116. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281. https://doi.org/10.1038/ncb3493.

117. Drissen, R., Thongjuea, S., Theilgaard-Mönch, K., and Nerlov, C. (2019). Identification of two distinct pathways of human myelopoiesis. Sci. Immunol. *4*, eaau7148. https://doi.org/10.1126/sciimmunol.aau7148.

118. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309. https://doi.org/10.1016/j.cell.2011.01.004.

119. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler [version 2; peer review: 2 approved]. F1000Research *9*. https://doi.org/10.12688/f1000research.24956.2.

120. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978. https://doi.org/10.1038/nmeth.4401.

121. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

122. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.

123. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-human CD45RA (clone HI100)	BD	Cat#564552; RRID:AB_2738841
Mouse anti-human CD90 (clone 5E10)	BioLegend	Cat# 328124; RRID:AB_2561693
Mouse anti-human CD123 (clone 6H6)	BioLegend	Cat# 306006; RRID:AB_314580
Rat anti-human CD49f (clone GoH3)	BioLegend	Cat# 313626; RRID:AB_2616782
Mouse anti-human CD38 (clone HIT2)	BioLegend	Cat# 303526; RRID:AB_10983072
Mouse anti-human CD10 (clone HI10a)	BioLegend	Cat# 312222; RRID:AB_2562157
Mouse anti-human CD34 (clone 581)	BioLegend	Cat# 343510; RRID:AB_1877153
Mouse anti-human CD117 (clone 104D2)	BioLegend	Cat# 313238; RRID:AB_2629837
Mouse anti-human CD2 (clone RPA-2.10)	BioLegend	Cat# 300210; RRID:AB_314034
Mouse anti-human CD3 (clone HIT3a)	BioLegend	Cat# 300310; RRID:AB_314046
Mouse anti-human CD4 (clone RPA-T4)	BioLegend	Cat# 300510; RRID:AB_314078
Mouse anti-human CD8a (clone RPA-T8)	BioLegend	Cat# 301010; RRID:AB_314128
Mouse anti-human CD11b (clone ICRF44)	BioLegend	Cat# 301308; RRID:AB_314159
Mouse anti-human CD14 (clone 61D3)	eBioscience	Cat#15-0149-42; RRID:AB_2573058
Mouse anti-human CD19 (clone HIB19)	BioLegend	Cat# 302210; RRID:AB_314240
Mouse anti-human CD20 (clone 2H7)	BioLegend	Cat# 302308; RRID:AB_314256
Mouse anti-human CD56 (clone MEM188)	BioLegend	Cat# 304608; RRID:AB_314450
Mouse anti-human CD235ab (clone HIR2)	BioLegend	Cat# 306606; RRID:AB_314623
Chemicals, peptides, and recombinant proteins		
RPMI-1640	Gibco	Cat#21875034
Fetal Bovine Serum	StemCell Technologies	Cat#06950
Penicillin-Streptomycin (10,000 U/mL)	Gibco	Cat#15140122
MEM α, nucleosides	Gibco	Cat#12571063
IMDM	Gibco	Cat# 21056023
DPBS	ThermoFisher	Cat# 14190169
30% (wt/vol) BSA	Sigma-Aldrich	Cat#A8577
Digitonin	Promega	Cat# G9441
IGEPAL CA-630	Sigma-Aldrich	Cat# I8896
Tween 20	Sigma-Aldrich	Cat# 11332465001
Tris-HCl 1M pH 8.0	ThermoFisher	Cat# 15568025
NaCl 5M	Sigma-Aldrich	Cat# S6546
MgCl2 1M	ThermoFisher	Cat# AM9530G
N,N-Dimethylformamide	Sigma-Aldrich	Cat# D4551-250ML
Magnesium acetate solution 1M	Sigma-Aldrich	Cat# 63052-100ML
Potassium acetate solution 5M	Sigma-Aldrich	Cat# 95843-100ML-F
Tn5	Illumina	Cat# 20034197
EDTA 0.5M pH 8.0	ThermoFisher Scientific	Cat# AM9260G
SDS 10%	Sigma-Aldrich	Cat# L4509
DAPI	ThermoFisher Scientific	Cat# 62248
UltraPure Agarose	Invitrogen	Cat# 16500-500
Ethidium Bromide Solution	Invitrogen	Cat# 15585-011
2x NEB High Fidelity Q5 master-mix	New England Biolabs	Cat# M0515
Nuclease-free water	Invitrogen	Cat# AM9932
PB buffer	Qiagen	Cat# 19066
Sodium acetate	ThermoFisher	Cat# AM9740

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
KAPA 2G Robust HS Ready Mix	Sigma-Aldrich	Cat# KK5702
FastStart High Fidelity polymerase	Sigma-Aldrich	Cat#4738292001
Access Array Barcode Library for Illumina Sequencers-384, Single Direction	Fluidigm	Cat# 100-4876
KAPA dual-indexed adapters	Roche	Cat# KK8722
KAPA HiFi HotStart high-fidelity DNA polymerase	Roche	Cat#7958927001
SeqCap EZ Prime Choice (custom probes)	Roche	N/A
Critical commercial assays		
MycoAlert Mycoplasma Detection Kit	Lonza	Cat#LT07-218
DNeasy Blood & Tissue Kit	Qiagen	Cat#69506
KAPA Hyper-Plus Kit (96)	Roche	Cat#KK8514
Qiaquick spin column	Qiagen	Cat# 28104
Qubit dsDNA HS Assay Kit	Invitrogen	Cat# Q32854
Agilent Tapestation HS D5000 ScreenTape	Agilent	Cat#5067-5592
Agilent Tapestation HS D5000 Reagents	Agilent	Cat#5067-5593
Agilent Tapestation HS D1000 ScreenTape	Agilent	Cat# 5067-5583
Agilent Tapestation HS D1000 Reagents	Agilent	Cat#5067-5584
Agilent High Sensitivity DNA Kit	Agilent	Cat#5067-4626
Deposited data		
Raw and processed scATAC-seq data generated with GTAC (cell lines and primary samples)	This study	GEO: GSE214812
Targeted single cell genotyping sequencing data generated with GTAC (cell lines and primary samples)	This study	SRA: PRJNA887752
Experimental models: Cell lines		
K562	Vyas lab	N/A
JURKAT	Vyas lab	N/A
OCI-AML3	Milne lab	N/A
THP1	Vyas lab	N/A
MOLM-13	Vyas lab	N/A
Oligonucleotides		
See Table S4 for target-specific genotyping oligos	IDT (design: this paper)	N/A
See Table S4 for target-specific nested barcoded genotyping oligos	IDT (design: this paper)	N/A
384 indexed i7 primers	IDT (design: Xu et al.[34])	https://www.nature.com/articles/s41596-021-00583-5
See Table S4 for Custom barcoded i5 primers	Illumina	N/A
CS1 sequencing primer (HPLC purified): A + CA+CTG+ACGACATGGTTCTACA	Rodriguez-Meira et al.[107]	https://www.sciencedirect.com/science/article/pii/S266616672030112X
CS2 sequencing primer (HPLC purified): T + AC+GGT+AGCAGAGACTTGGTCT	Rodriguez-Meira et al.[107]	https://www.sciencedirect.com/science/article/pii/S266616672030112X
CS2rc sequencing primer (HPLC purified): A+GAC+CA+AGTCTCTGCTACCGTA	Rodriguez-Meira et al.[107]	https://www.sciencedirect.com/science/article/pii/S266616672030112X
Software and algorithms		
Burrows-Wheeler aligner 0.7.17	Li et al.[103]	https://academic.oup.com/bioinformatics/article/25/14/1754/225615
VarDictJava 1.7.0	Lai et al.[104]	https://academic.oup.com/nar/article/44/11/e108/2468301
Annovar	Yang et al.[106]	https://www.nature.com/articles/nprot.2015.105
Primer-BLAST	Ye et al.[108]	https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-134

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TARGET-seq genotyping pipeline	Rodriguez-Meira et al.[107]	https://github.com/albarmeira/TARGET-seq
Bcl2fastq	Illumina	https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf
Snakemake scATAC-seq pre-processing pipeline	Chen et al.[33]	https://github.com/dbrg77/scATAC_snakemake
LanceOtron peak caller	Hentges et al.[111]	https://academic.oup.com/bioinformatics/article/38/18/4255/6648462
R 4.1.3	R-Project	
ArchR 1.0.2	Granja et al.[112]	https://www.nature.com/articles/s41588-021-00790-6
chromVAR 1.16.0	Schep et al.[120]	https://www.nature.com/articles/nmeth.4401
GraphPad Prism 9	GraphPad	https://www.graphpad.com
FlowJo	FlowJo	https://www.bdbiosciences.com/en-gb/products/software/flowjo-v10-software
g:Profiler	Kolberg et al.[119]	https://f1000research.com/articles/9-709/v2
Fgsea	Korotkevich et al.[121]	https://www.biorxiv.org/content/10.1101/060012v3
infSCITE	Jahn et al.[110]	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0936-x
BioRender	BioRender	https://biorender.com
Mutect2	Cibulskis et al.[105]	https://www.nature.com/articles/nbt.2514
Custom code for data analysis	This study	https://doi.org/10.5281/zenodo.7817225
Other		
12.5 mL GRIPTIP, Sterile, Filter	INTEGRA Biosciences	Cat#6455
High Volume MANTIS Chip	FORMULATRIX	Cat# MCHVSMR6
Agencourt AMPure XP beads	Beckman Coulter	Cat# A63881

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Paresh Vyas (paresh.vyas@imm.ox.ac.uk).

Materials availability

The list of all oligo sequences designed in this study and used for single-cell genotyping can be found in Table S4. These include both target-specific oligos used in the first PCR, and nested barcoded target-specific oligos used in genotyping PCR1.

Data and code availability

Raw and processed scATAC-seq data have been deposited at GEO and are publicly available as of the date of publication. Raw targeted single-cell genotyping data produced with GTAC have been deposited at NCBI's SRA data repository and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Python and Perl scripts used for scATAC or single-cell genotyping data pre-processing were published previously[17,33] and are available at the respective Github pages. All original code has been deposited at Github and is publicly available as of the date of publication. DOIs are listed in the key resources table. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

K562, MOLM-13, THP1 and JURKAT human cell lines were cultured in RPMI-1640 medium (Cat# 21875034, Gibco) supplemented with 10% FBS and 1% V/V Pen-Strep (Cat# 15140122, Gibco). OCI-AML3 cells were cultured in Alpha-MEM medium with nucleosides (Cat# 12571063, Gibco) supplemented with 20% FBS and 1% V/V Pen-Strep. All cell lines were regularly screened for Mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Cat# LT07-218). Cells were passaged every 2–3 days and seeded at approximately 500 000 cells/mL. All cell lines were kept in a CO_2 incubator at 37°C. Prior to GTAC procedures, viability was evaluated to assure >90% viability.

Patient samples

Patient BM samples were collected with informed consent under ethically approved protocols (MREC 06/Q1606/110 or NHS REC 17/YH/0382). Written informed consent was obtained in accordance with the Declaration of Helsinki. Cytogenetic analyses were performed in clinical laboratories. Mononuclear cells (MNCs) were isolated by Ficoll density gradient centrifugation and viably frozen in 90% FBS/10% DMSO. Genomic DNA was extracted using a DNeasy Blood & Tissue Kit (Cat# 69506, Qiagen). The diagnostic human AML BM sample used for GTAC experiments was derived from a male AML patient aged 81 at the time of sample collection. Wild-type BM used for GTAC experiments was derived from a female orthopedic patient undergoing hip replacement surgery, aged 65 at the time of sample collection.

METHOD DETAILS

Targeted DNA sequencing

Library preparation and sequencing

Targeted DNA sequencing was performed on bone marrow MNCs. Pre-capture DNA libraries were prepared using the KAPA HyperPlus protocol (Cat# KK8514, Roche). 100ng of genomic DNA was fragmented by enzymatic fragmentation. Following end repair and A-tailing, adapter ligation was performed using KAPA dual-indexed adapters (Cat# KK8722, Roche). Library cleanup and double-sided size selection was performed using Agencourt AMPure XP beads (Cat# A63881, Beckman Coulter) to obtain fragments of ~320 bp. Libraries were amplified by ligation-mediated PCR for 6 cycles using a KAPA HiFi HotStart high-fidelity DNA polymerase (Cat# 7958927001, Roche) and purified using AMPure XP beads.

Targeted capture was performed using a custom pool of biotinylated capture probes (SeqCap EZ Prime Choice, Roche) targeting 97 genes recurrently mutated in myeloid malignancies spanning 347 kb.[2,16,99–102] Amplified DNA libraries were hybridized to the capture probes in pools of 10–12 samples according to the manufacturer's instructions. The captured DNA libraries were amplified by 14 cycles of PCR using a KAPA HiFi HotStart DNA polymerase and purified using AMPure XP Beads.

Post-capture amplified DNA libraries were quantified by Qubit (Life Technologies) and size distribution and quality analyzed using a Bioanalyser chip (Agilent Technologies). Libraries were pooled in equimolar concentrations and were sequenced on an Illumina NextSeq 500 using paired-end reads.

Variant calling in targeted panel sequencing

Sequencing data were analyzed in a custom pipeline based on the GATK Best Practices (GATK v4.1.2.0 and Picard v2.21.3). Raw sequencing reads were converted to an unmapped BAM file and adapter sequences soft-clipped using Picard *MarkIlluminaAdapters*. Following conversion back to a FASTQ file, reads were mapped to the hg38 human reference genome assembly using the Burrows–Wheeler aligner v0.7.17[103] with alternate contig-aware alignment. Mapped, and unmapped, BAM files were merged using *MergeBamAlignment* and reads from different sequencing lanes were combined. Duplicate reads were marked using Picard *MarkDuplicates* and base quality scores recalibrated with GATK BaseRecalibrator and ApplyBQSR. Somatic variant calling was performed on the pre-processed BAM files using VarDictJava v1.7.0[104] and Mutect2[105] in tumour-only mode. For VarDict, variants were called with a minimum variant allele frequency of 0.01, minimum base quality score of 25 and minimum supporting reads of 2, with indel realignment and removal of adapter sequences. For Mutect2, a minimum tumor LOD of 2 was used and variants were filtered for sequence context-dependent artifacts using *FilterMutectCalls* and *FilterByOrientationBias*. Indels were left-aligned and normalized using bcftools norm (v1.9). Variants were annotated using Annovar.[106] Target enrichment metrics and coverage was calculated using Picard *CollectHsMetrics* and custom scripts.

VarDict and Mutect2 variant calls were analyzed separately to identify a consensus list of high-confidence variants. The following post-processing filters were applied to VarDict calls to exclude likely sequencing artifacts: (1) Minimum of 5 variant reads for SNVs (with at least 2 reads in forward and reverse directions), or minimum of 10 variant reads for indels (with at least 4 reads in forward and reverse directions). (2) Minimum base quality score 30. (3) Minimum mapping quality score 40, except for variants in *U2AF1;U2AF1L5*, where the mapping quality was ignored. This because, there is a duplication of the *U2AF1* gene on chromosome 21 called U2AF1L5, in hg38, which results in reads being flagged as multi-mapped. (4) Maximum strand bias Fisher p value of 0.0001. (5) No position bias toward beginning or end of reads. The following post-processing filters were applied to Mutect2 calls: (1) Passed all default Mutect2 filters or only failing the *clustered_events* filter. (2) Minimum of 5 variant reads for SNVs (with at least 2 reads in forward and reverse directions), or minimum of 10 variant reads for indels (with at least 4 reads in forward and reverse directions).

Variants were flagged as likely germline, or sequencing artifacts, if any of the following applied: (1) Variant allele with a population allele frequency >1 in 1,000 according to any of three large polymorphism databases (Gnomad, 1000 Genomes Project, ESP6500) that is not a hotspot driver mutation with a COSMIC (v88) occurrence count of >100 cases. (2) Present in a panel of normal cord blood samples. (3) Recurrent in the cohort unless present at least 5 times in COSMIC.

Primary bone marrow cell sorting

Thaw media was prepared with IMDM buffer (Cat# 21056023, Gibco) supplemented with 20% FBS and 11% DNAse I (110 μg/mL). Cryopreserved bone marrow samples were thawed at 37°C in a water bath, 1 mL warm FBS was added and the suspension then diluted by dropwise addition of 8 mL thaw media. Cells were centrifuged at 350 g for 10 min at room temperature and then resuspended in FACS buffer (IMDM with 10% FBS and 10 μg/mL DNAse I), filtered, and placed on ice. Cells were stained for flow cytometry with antibodies listed in key resources table for 20–40 min on ice. After washing, cells were centrifuged at 350 g for 5 min at 8°C,

resuspended in 200–500 μL FACS buffer containing DAPI at 1 in 10^4 dilution for dead cell exclusion. Cells were sorted on a Sony MA900 instrument. Single stains were used to set initial voltages. Fluorescence minus one controls (FMOs) and positive and negative cell populations were used to set gates. Single, live $CD34^+$ cells were sorted into IMDM with 10% FBS, in tubes that were pre-coated with IMDM containing 10% FBS.

GTAC cell lysis and tagmentation

Tagmentation of cell populations was performed using the Omni-ATAC protocol.[22,34] Briefly, 50,000–90,000 FACS-sorted cells were centrifuged, washed with 500 μL of cold PBS containing 0.5% BSA, and centrifuged again at 500 g in a pre-chilled (4°C) fixed-angle centrifuge for 5 min. The supernatant was carefully removed. Cell pellets were then resuspended in 50 μL ATAC resuspension-digitonin buffer (ATAC-RSB-DTN buffer; 0.01% digitonin, 0.1% NP-40, 0.1% Tween 20, 10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 3 mM MgCl2) and left on ice for 3 min. Following lysis, 1 mL chilled RSB-Wash buffer (0.1% Tween 20, 10 mM Tris-HCl pH 8.0, 10 mM NaCl, and 3 mM MgCl2) was added and the tube was inverted gently a few times. Nuclei were centrifuged at 1000 g for 3–5 min at 4°C and the supernatant removed leaving 2–3 μL of buffer. Nuclei were resuspended in chilled tagmentation buffer (16% DMF, 66mM potassium acetate, 10mM magnesium acetate, 33mM Tris-HCl pH 8.0, 0.1% digitonin) containing 5 μL of Tn5 (Cat# 20034197, Illumina). Tubes were incubated in a pre-heated thermo-mixer for 30 min at 37°C with 800 rpm shaking. After incubation, 50 μL tagmentation stop buffer containing 10 mM EDTA and 0.5% BSA was added, the suspension was pipette-mixed gently 10 times and left on ice for a minimum of 10 min.

GTAC lysis plate preparation, single-nucleus sorting and preamplification

All cell line experiments were done in the 96-well plate format, while primary samples were processed using a 384-well plate format. Here we describe the procedure for 384-well plates. All the volumes were doubled for the 96-well format, as recommended previously.[34] Note that all the experiments prior to the first PCR (lysis buffer preparation, plate pre-processing after sort) were carried under the hood in a designated pre-PCR clean room. The full list of the 384 barcoded i7 indexes have been published.[34] Targeted oligo sequences used for genotyping and barcoded nested target-specific primers used for PCR1 are listed in Table S4.

1.1X lysis buffer was prepared as detailed previously.[34] 27 μL of 1.1X lysis buffer was aliquoted into each well of a 384-well plate (Cat# 4ti-0384/B, FrameStar) and 3 μL of 100 μM i7 indexed primers (containing well barcodes; IDT) were added to each well to obtain a 10μM i7 index stock plate with unique barcodes in each well. Sorting plates were prepared by transferring 1 μL from the 10μM i7 index stock plate into a new 384-well plate and adding 1 μL of 10 μM i5 indexed primer (containing a plate barcode; IDT) in 1X lysis buffer, using a unique i5 index for each plate. Sorting plates were stored at −80°C and thawed prior to nuclei sorting.

Nuclei were sorted on a Sony MA900 instrument. 500–1500 μL of nucleus dilution buffer (0.5% BSA in 0.5X PBS and nuclease-free water) was added to the tagmented nuclei suspension and 1:1000 1 mg/mL DAPI was added for a final DAPI dilution $1:10^5$ (an unstained sample was used to set the DAPI-negative gate). Single nuclei were identified using a combination of FSC, SSC and DAPI fluorescence and sorted into each well of the 384-well plates, excluding clumps and doublets with two consecutive singlet gates. Plates were stored at −80°C.

Next, plates were incubated at 65°C for 15 min followed by addition of 1 μL Tween 20 10% (Cat# 11332465001, Sigma). PCR preamplification was performed by adding 5 μL PCR mix containing 2x NEB High Fidelity Q5 master-mix (Cat# M0515, New England Biolabs) and target-specific genotyping primers (IDT) at a final concentration of 100 nM per primer. Plates were incubated on a thermo-cycler with the following PCR conditions: 72°C for 10 min, 98°C for 1 min, then 15–18 cycles of 98°C 10 s, 60°C 30 s and 72°C for 30 s, then a final elongation at 72°C for 5 min 18 cycles of amplification were used for cell lines and 16 cycles of amplification for primary human cells.

After the preamplification PCR, 20% of the PCR product was transferred to a new plate containing 4 μL nuclease-free water (Cat# AM9932, Invitrogen) per well to make a stock plate for single-cell genotyping which was then stored at −80°C. To purify scATAC libraries, the remaining 80% of the volume from each 384-well plate was pooled into a Falcon tube. Each pool was mixed with 5 volumes of Qiagen PB buffer (Cat# 19066, Qiagen) and 10–30 μL of sodium acetate pH 5.5 (Cat# AM9740, ThermoFisher) was added. The pool was purified using a QIAquick spin column (Cat# 28104, Qiagen) and eluted into a total of 42 μL of nuclease-free water (two elutions in 21 μL were done). The pooled libraries were further purified twice with AMPure XP beads first with a 0.5:1 ratio of beads to sample, then with a 1.2:1 ratio of beads to sample. Libraries were quantified with Qubit dsDNA HS Assay Kit (Cat# Q32854, Invitrogen) and fragment size was analyzed with Agilent Bioanalyzer (High Sensitivity DNA Kit, Agilent, Cat# 5067-4626) or Agilent Tapestation (High Sensitivity D5000 ScreenTape and Reagents, Agilent, Cat# 5067–5592 and Cat# 5067–5593). Libraries were pooled in equimolar concentrations and diluted to 4 nM for sequencing.

GTAC single-nucleus genotyping library preparation

To generate Illumina-compatible libraries for single-cell genotyping, two PCR steps are required following the initial preamplification PCR. In the first PCR step (genotyping PCR1), nested target-specific primers containing universal CS1 (forward primer) or CS2 (reverse primer) adapters are used to amplify the target loci of interest. Incorporation of a barcode sequence specific to each plate into these primers enables libraries from different plates to be pooled later (Figure 1). Genotyping stock plates with the preamplification PCR product were thawed and spun at 3000 g at room temperature. PCR1 reactions were set up using 3.25 μL of KAPA 2G Robust HS Ready Mix (Cat# KK5702, Sigma-Aldrich), 1.5 μL of preamplification PCR product and 300 nM target-specific primers containing plate barcodes (IDT) in a 6.5 μL reaction. PCR1 was performed with the following PCR conditions: 95°C for 3 min, 22–25

cycles of 95°C 15 s, 60°C 20 s and 72°C for 1 min, then a final elongation at 72°C for 5 min. For cell lines, 22 cycles of PCR were used and for primary human cells, 24 cycles were used.

In the second PCR step (genotyping PCR2), Illumina-compatible adapters containing a 10 bp cell barcode are attached to the genotyping PCR1 product by binding to the CS1/CS2 adapters. PCR2 reactions were performed using FastStart High Fidelity polymerase (Sigma-Aldrich) with 1.0 μL of PCR1 product and 1.2 μL of each barcode primer mix (Access Array Barcode Library for Illumina Sequencers- 384, Single Direction, Cat# 100–4876, Fluidigm) in a 6.2 μL reaction. PCR2 was performed with the following program: 95°C for 10 min, then 10 cycles of 95°C 15 s, 60°C 30 s and 72°C for 1 min, then a final elongation at 72°C for 5 min.

After PCR2, 1 μL from each well was pooled using Mosquito nanolitre liquid handling platform (TTP Labtech) and purified with AMPure XP beads with a 1.2:1 ratio of beads to PCR product. Libraries were quantified with Qubit dsDNA HS Assay Kit and fragment size was analyzed with Agilent Tapestation (High Sensitivity D1000 ScreenTape and Reagents, Agilent, Cat# 5067–5583 and Cat# 5067–5584). Next, libraries were pooled at equimolar ratios taking into consideration average library size of the library, total number of amplicons per pool (number of amplicons per cell * number of cells) and Qubit measurements. The pool underwent a final AMPure XP bead purification with a double-size selection, using first a 0.6:1 ratio of beads to library, followed by a 1.2:1 ratio of beads to library to remove traces of scATAC library and preserve only genotyping amplicons. The final library was diluted to 4 nM prior to sequencing.

GTAC sequencing

For scATAC library sequencing, the number of reads per cell depends on the library complexity of a given sample. From our experience, at least 200k reads per cell are recommended. K562 cells used for scATAC library comparison (Figure 3) were sequenced at average depth of 435080 reads/cell. The AML sample was sequenced at average depth of 256064 reads/cell. For scATAC library sequencing, we either used Illumina NextSeq 500/550 Mid-Output kits or NextSeq 500/550 High-Output kits. All sequencing runs were performed using 150-cycle kits (75 bp paired-end).

For genotyping libraries, we aimed for a minimum of 2000 reads per cell per amplicon. We used either Illumina MiSeq v3 or NextSeq Mid Output kits (75 bp paired-end). Three custom primers were used for sequencing: two primers anneal to the CS1 and the CS2 adaptors and sequence the plate barcode and the specific region of interest (500 nM CS1 primer and 500 nM CS2 primer), while the CS2rc primer anneals to the reverse complement of the CS2 sequence, reading the 10-bp Fluidigm well barcode (500 nM CS2rc primer).[107] All sequencing primers were diluted in 700 mL of HT1 buffer.

GTAC primer design and validation

Target-specific preamplification primer pairs were designed using Primer-BLAST,[108] using the *Homo sapiens* database, using *Thermodynamic Oligo Alignment* option. Ideally, amplicons should be 250 bp or shorter in length, given that longer regions might fail amplification due to Tn5 tagmentation (if the region is in open chromatin). Primers were checked with the IDT Oligo Analyzer tool to exclude annealing to i7 and i5 ATAC primers, avoiding those with $\Delta G < -9$ kcal/mol. Nested target-specific primers used in genotyping PCR1 were also designed with Primer-BLAST and additionally contain universal CS adaptors (CS1 adaptor for forward nested primer and CS2 adaptor for the reverse nested primers) followed by a 6 bp-long custom plate barcode. We note that when genotyping libraries are sequenced with 75 bp paired-end sequencing, at least one of the nested primers should not be more than 69 bp from the locus of interest. Primer pairs were first tested using extracted genomic DNA and appropriate enzymes. Pairs displaying clear bands of expected sizes and low or no unspecific amplification were tested in single-cell reactions. For that purpose, single OCI-AML3 or primary $CD34^+$ bone marrow cells were sorted into lysis buffer containing generic i5 and i7 oligos and GTAC library preparation was carried out as described in the previous sections. Each primer pair was tested on 16 single cells. After 18 cycles of PCR, an aliquot of the sample was taken for genotyping testing, in which the PCR1 reaction was performed for each single cell with 32 cycles of amplification, followed by amplicon evaluation on a 2% agarose gel. To confirm correct scATAC library formation, the remaining material from each set of 16 cells was pooled and purified as described above, followed by fragment size analysis on an Agilent Bioanalyzer. Primer pairs forming primer dimers or interfering with correct library formation were excluded and re-designed.

Single-cell genotyping data pre-processing and mutation calling

Single-cell genotyping data was pre-processed using the published custom pipeline used previously for TARGET-seq (https://github.com/albarmeira/TARGET-seq).[107] Reads were first demultiplexed using the 384 well barcodes introduced via the genotyping PCR2 reaction, followed by demultiplexing based on plate barcodes introduced during genotyping PCR1. This generated separate fastq files for each single cell. Reads were aligned to hg38 using STAR version 2.7.3a with default settings. Variant calling was performed with *mpileup* (samtools version 1.1, options: *–minBQ 30*, *–count-orphans*, *–ignore overlaps*) and results were summarized using the custom pipeline. This generates summary files for each variant, providing the number of A, G, T, C or indel reads for the specified coordinates for each cell. We note that in case of indels, it might be necessary to provide the coordinates mapping 1–2 bp downstream of the first affected base, in order to avoid the indel being missed by the variant call. For FLT3-ITD genotype pre-processing (MOLM-13 cell line), getITD[109] pipeline was used instead, providing the primer and CS1/2 adaptor sequences, and using *-min_read_length 60*.

Mutational calling on single cells was then performed with custom R scripts, separately for each variant. Coverage for each cell was calculated as the sum of all reads across the variant locus for that cell. Empty wells routinely displayed zero or very few reads (usually

up to 2), indicating no cross-well contamination. Cells were filtered out if the coverage was less than 30 reads across the variant locus. The scVAF for each cell was calculated with the following formula:

$$\text{scVAF} = \text{variant reads} / \text{coverage}$$

PCR amplification and next-generation sequencing have an inherent error rate, which needs to be accounted for when calling cell genotypes. Control cells that were WT for the locus were used to set scVAF thresholds for mutation calling. For mutational analysis in cell lines, we used cell line(s) that were WT for the locus as a control. For mutational analysis of the human AML sample, we profiled 300 $CD34^+$ bone marrow cells from a healthy individual WT for the loci of interest. The scVAF threshold for calling a cell mutant was set using the following formula:

$$\text{VAF threshold} = \max(\text{scVAF in WT control}) + 2 \times \text{SD of scVAF in WT control}$$

Furthermore, we set a threshold for the minimum number of mutant reads for a cell to be called mutant, calculated with the following formula:

$$\text{Minimum number mutant reads} = \text{Maximum number mutant reads in WT sample} + 1$$

Thus, cells were called mutant for a locus if both the scVAF threshold and the minimum number of mutant reads were passed.

Inference of clonal hierarchies

For the AML sample, each single cell was assessed for detection or non-detection of the five mutations present in the sample. The pattern of mutational co-occurrence was used to determine the most likely clonal structure and assign a clonal identity to each cell using principles based on those in our previous study.[8] The occurrence of ADO means that in some cells, a mutation that is present will not be detected. For most mutations in this sample, the sequence of acquisition was clear. However, we observed a small number of cells belonging to TA and TR clones, suggesting either the *ASXL1* or the *RUNX1* mutation was acquired after *TET2*. We observed 21 cells with the TA genotype compared to 11 cells with the TR genotype, suggesting that T → TA → TAR was the most likely sequence of acquisition.

While our highly sensitive single-cell genotyping resulted in straightforward clone assignment for most cells, a fraction of cells failed detection or displayed ADO for one or more loci. In some of these cases, we were still able to assign a cell to a clone. For example, a small proportion of cells were mutant for *TET2* and one of the *BCOR* loci but the *ASXL1* or *RUNX1* mutations were not detected. In these cases, we inferred that ADO had occurred, and the cell was assigned to the appropriate daughter clone. We applied the same reasoning in cells that were called WT for *TET2* but mutant for *ASXL1*, *RUNX1* and/or one of the *BCOR* loci, in which we inferred ADO of the mutant *TET2* allele had occurred. In all cases, this was a rare occurrence consistent with our estimates of the ADO rate.

We could not apply the same ADO correction for loci harboring later mutations (*BCOR* mutations), given the absence of a later 'control' event. In a fraction of cells, mutations were called on *TET2*, *ASXL1* and *RUNX1*, while one *BCOR* locus was called WT and the other failed detection (less than 30 reads covering the variant allele). In those cases, we assigned the cell to the TAR clone. If mutations were called on *TET2*, *ASXL1* and *RUNX1* but both *BCOR* loci failed detection, we removed these cells from further analysis. We also noticed 18 cells where both *BCOR* mutations were detected, and scVAFs were consistent with heterozygosity, suggesting they were doublets. Consistent with this, the majority of these cells had a high number of unique scATAC fragments. We therefore excluded these from downstream analysis.

To confirm that our phylogenetic tree for the AML sample represented the statistically most likely course of somatic events, we used infSCITE[110] with default parameters and '-r 200 -L 10000 -fd 0.01 -ad 0.02 -e 0.2 -p 1000'. As input, we used the matrix containing the mutational status for each locus, for each analyzed cell. The estimated allelic drop-out rate determined by infSCITE was 2%.

Allelic dropout (ADO) estimation

To estimate ADO rates for a range of loci, we analyzed variants known to be present in a heterozygous state in all cells and calculated the frequency with which they were called homozygous (either reference or alternative) when they should be heterozygous.

For the cell line analysis, we profiled 5 heterozygous loci in K562 cells (*HOXA9*, *CBL*, *NOTCH1*, *ASXL1*, *TET2*), 2 in OCI-AML3 (NPM1, DNMT3A), 1 in MOLM-13 (FLT3) and 1 in JURKAT (*RUNX1*). In the primary AML sample, we profiled two germline heterozygous SNPs (in the *SH2B3* and *ASXL2* loci) (Figures S2C–S2E), which were previously identified via targeted DNA sequencing of BM MNCs. For each locus, detection of the variant allele was called by setting a scVAF threshold based on WT control cells as described previously:

scVAF threshold for detection of variant allele = max scVAF in WT + 2 x SD of the scVAF mean in WT

To determine detection of the reference (or WT) allele, we defined a scVAF threshold using the following formula:

$$\text{scVAF threshold for detection of reference allele} = 1 - \text{scVAF threshold for detection of variant allele}$$

For each allele, we then calculated the percentage of cells in which ADO occurred:

$$\text{ADO percentage} = (\text{Cells below scVAF threshold}) / (\text{Total number of cells}) * 100$$

Finally, total ADO rate was calculated as:

$$\text{Total ADO percentage} = \text{Variant ADO percentage} + \text{WT ADO percentage}$$

We therefore estimate that the percentage of cells in which biallelic detection occurred as:
Biallelic detection percentage = 100 – Total ADO percentage
We report variant ADO, WT ADO and biallelic detection percentages in Results. We note that, when profiling heterozygous mutations at single-cell level, WT ADO is not usually problematic because the cell can still be called mutant. The same is not true for mutant ADO, which leads to a mutant cell being wrongly assigned as WT.

Coverage calculation from scATAC data

To calculate the number of reads mapping to target loci of interest within scATAC libraries, we used *bedtools coverage* on non-de-duplicated BAM files for each scATAClibrary, providing a BED file containing variant coordinates. We then plotted the coverage as total reads mapping to the locus derived from each cell.

scATAC data pre-processing

Raw bcl files were first demultiplexed using *bcl2fastq* (version 2.20) and specifying - - *no-lane-splitting*. Files were demultiplexed using unique i7/i5 combinations present in each well. Raw data was pre-processed with a published pipeline written in snakemake,[33] with minor modifications. In brief, trimming was performed with cutadapt42 to remove the Nextera sequence at the 3′-insert end. Next, reads were mapped to hg38 with Hisat2. Reads with mapping quality less than 30 were removed by samtools (*-q 30* flag). Next, each bam file was de-duplicated using the Picard *MarkDuplicates* function and indexed using *samtools index*. Files with aggregate size distribution for each cell were generated using samtools after removal of the reads mapping to chrM. De-duplicated reads from single cells were merged using samtools (*ulimit -n 16384* was specified to avoid errors if more than 1024 cells were analyzed). Average fragment size was computed again on the merged file excluding chrM. The merged BAM file was converted to BED format via *bam2bed*. The BED file was used for peak calling with MACS2 and specifying *–nomodel –shift -100 –extsize 200*. The resulting BDG files were used to generate a genome coverage file in BW format using *bdg2bw,* which was used as input for UCSC genome browser visualization, as well as for LanceOtron peak caller[111] used to generate peak sets for correlation analyses (Figure 3F, detailed below). A count matrix over the union of peaks was generated by counting the number of reads from individual cells that overlap the union peaks using *coverageBed* from bedTools. Additional per-cell quality metrics were generated from individual BAM files using MACS2-produced.narrowPeak files for peak coordinates.

Correlation of peak calls

Merged BAM files were indexed using *samtools index* function. Then, coverage files in BW format were produced using *bamCoverage - - extendReads -bs 1 - - normalizeUsing RPKM*. Coverage files were loaded into the LanceOtron peak caller in order to extract bona-fide peaks of open chromatin and filter out noise peaks to avoid inflating correlation (model used: wide-and-deep-jan-2021). Output files were filtered for true peaks via peak scores – a bona-fide peak of open chromatin had the peak score higher than 0.5. Next, these files were merged to create a union of peak coordinates sorted by chromosome and start coordinates. The final union file was generated with *bedtools merge,* containing all bona-fide peaks called across all samples. *multiBigwigSummary* quantified sample correlation across peaks present in the union bed file, taking as input original coverage files, excluding chrM peaks. To plot correlation results, *plotCorrelation* was used with the following parameters: - - *corData results.npz - - corMethod pearson - - whatToPlot heatmap - - removeOutliers - - skipZeros - - plotNumbers*.

Merging scATAC and genotyping data

To link clonal information with ATAC-seq data for each single cell, genotyping data was merged with metadata produced by the pre-processing pipeline using shared cell IDs. ArchR[112] was used for most downstream scATAC analyses (details in the next section). To generate Arrow Files and the ArchR project, the fragment file produced by the pre-processing pipeline was used as input. Arrow Files were generated using *maxFrags = 120000*, *filterTSS = 8* and *filterFrags = 10000* and *addGeneScoreMat = TRUE* to add gene score information used for later analyses. After ArchR project generation, the relevant metadata was extracted using *cellColData*. Next, this table was merged with the pre-processing metadata via the cellNames column to link clonal information with the QC metrics assigned during Arrow File generation, maintaining the row order from the ArchR metadata, and subtracting cells filtered out during Arrow File generation. The column containing clonal information was then extracted in vector format and added as new column to the ArchR project metadata. Finally, cells which could not be assigned to a clone were removed from the ArchR project.

Doublet exclusion

We added doublet scores for each cell using the *addDoubletScores* function in ArchR, with *k = 10*, *knnMethod = 'UMAP'* and *LSIMethod = 1*. Briefly, this step generates synthetic doublets by using the provided data and by mixing sequencing reads derived from thousands of single-cell combinations. Synthetic doublets are then projected on a UMAP embedding and their closest neighbor is identified. The procedure is iterated thousands of times to identify most likely doublets within the data. Identified doublets were removed using *filterDoublets* with *cutEnrich = 1* and *filterRatio = 1*.

Dimensionality reduction and clustering

We performed Latent Semantic Indexing (LSI) dimensionality reduction using the *addIterativeLSI* function in ArchR, with *iterations = 2*, *varFeatures = 30000*, *dimsToUse = 1:30*, and *clusterParams* including *resolution = c(0.2)* and *sampleCells = 4000*. Clustering was performed using the *addClusters* function with the LSI dimensions as input using *method = 'Seurat'*, *resolution = 0.55*, and visualised by computing a UMAP using the *addUMAP* function with *nNeighbours = 30*, *minDist = 0.5*.

We used several orthogonal strategies to annotate clusters. First, we obtained marker genes for each cluster based on gene activity scores,[112] a measure of accessibility within and around gene body. Marker genes were identified using the *getMarkerFeatures* function with *useMatrix = 'GeneScoreMatrix', testMethod = 'wilcoxon', bias = c('TSSEnrichment', 'log10(nFrags)')*. This uses the Wilcoxon rank-sum test to identify marker genes while controlling for the TSS enrichment and $\log_{10}$(unique fragments) of cells when sampling the background set of cells. Marker genes with $\log_2$FC $\geq$ 0.25 and FDR $\leq$0.05 were used to test for enrichment of published human haematopoietic gene signatures[113–118] using g:Profiler[119] (Figure S3B), using 'Only annotated genes' as statistical domain scope. Second, we computed motif accessibility deviation scores of key transcription factors (Figure S3C). Third, we identified differentially accessible marker peaks for each cluster using the *getMarkerFeatures* function in ArchR (as described below) and identified the overlap of those marker peaks with curated peak sets from previously published bulk ATAC-seq data using the *peakAnnoEnrichment* function in ArchR (Figure S3D). Finally, we confirmed cluster annotations by projecting our data onto a healthy human bone marrow dataset[41] (details below, Figure S3E).

Gene activity and module score

We calculated gene activity scores in ArchR as previously described.[112] To plot gene activity scores on the UMAP embedding, MAGIC was used for signal smoothing across nearby cells to account for data sparsity, via *plotEmbedding* colored by the GeneScoreMatrix and using *imputeWeights = getImputeWeights*.

In order to compute the LSC score, we used the *addModuleScore* function in ArchR, providing as input the previously published LSC-specific signature,[43] and using gene activity scores from the GeneScoreMatrix.

Pseudo-bulk replicate generation, peak calling, and marker identification

We generated pseudo-bulk replicates using clonal information – each clone constituted a pseudo-bulk replicate. We used *addGroupCoverages* with *groupBy = 'clone'* and called peaks with MACS2 using *addReproduciblePeakSet* with *groupBy = 'clone'*, *shift = -75, extsize = 150*, *additionalParams = '--nomodel --nolambda'*, *peaksPerCell = 1000*, performing an iterative overlap procedure for merging peaks specific to ArchR. We called 175563 peaks in the AML sample. We added the peak matrix via *addPeakMatrix*.

Differential accessibility analysis and motif enrichment

We used several strategies to estimate chromatin accessibility differences between clones and clusters. We first defined marker peaks for clones and clusters using *getMarkerFeatures* with *useMatrix = 'PeakMatrix', bias = 'c('TSSEnrichment', 'log10(nFrags)'), testMethod = 'wilcoxon'*. Next, we extracted motif enrichment using *peakAnnoEnrichment* defining the *seMarker* object, with *peakAnnotation = 'Motif'*. Unless otherwise specified, we used the following cutoffs to select enriched motifs: FDR $\leq$0.05 and Log2FC $\geq$ 1. Heatmaps were generated using *plotEnrichHeatmap* and *ComplexHeatmap* functions. For motifs of TF belonging to TF families, we selected the TFs that were expressed in the cells where the motif was identified and where prior published evidence indicated a functional role. To obtain custom annotations for motif enrichment, we first added the annotation via *addArchRAnnotations*, specifying the *collection* parameter, and then added that through *peakAnnotation* within *peakAnnoEnrichment*.

For pairwise comparisons of motif enrichments between clones or clusters, we used deviations scores computed with chromVAR[120] within ArchR. We first added a background peak set via *addBgdPeaks* and then computed the deviation matrix using *addDeviationMatrix* with *peakAnnotation = 'Motif'*. For pairwise comparisons, we used *getMarkerFeatures* specifying *useGroups, bgdGroups, useMatrix = 'MotifMatrix', useSeqnames = 'z'* to compute differences in z-scores for motifs. To compare two different clones, we specified *groupBy = 'clone'* and clone names in *useGroups* and *bgdGroups*. We then extracted motifs with the respective FDR and 'MeanDiff' values to generate volcano plots. For pairwise comparison of peak accessibility, we used *getMarkerFeatures* specifying *useMatrix = 'PeakMatrix',* and utilized the information about the closest gene computed by ArchR. We similarly performed pairwise analyses between clones using gene activity scores; in this case, we used *getMarkerFeatures* specifying *useMatrix = 'GeneScoreMatrix'* and the groups to compare. For pathway enrichment analysis, we used the fgsea (v1.20.0) package.[121] We first used *getMarkerFeatures* to compare groups of interest followed by *getMarkers* with *cutOff = 'FDR <= 1'*. We then ranked the differentially accessible genes by multiplying the -$\log_{10}$(FDR) by the sign of the Log2FC (can be either 1 or -1). The ranked list was used within the *fgsea::fgseaMultilevel* function with *minSize = 5*, *maxSize = 1000*, and *nPermSimple = 10000*. We tested the ranked list against Hallmark, CGP, TFT, and GO human collections from MSigDB. The result was filtered only for pathways showing FDR <0.05.

Co-accessibility

We added loops indicative of peak co-accessibility using the *addCoAccessibility* function. Co-accessibility was generated together with the coverage tracks using the *getCoAccessibility* function, using *resolution = 1* and *corCutOff = 0.5*.

Projection analysis

To project the cell type label from healthy human hematopoiesis reference scATAC-seq data to our query AML dataset produced with GTAC, we firstly removed the unknown clusters, obtaining 33819 single-cells from the reference dataset.[41] We then converted our original fragment file to the hg19 annotation and repeated peak calling as described above. Then, peaks from both datasets were intersected, obtaining 162505 shared peaks by employing *GetIntersectingFeatures* in Signac (v1.9.0) package.[122] To work on the same genomic regions, peaks in the AML dataset were replaced with peak counts corresponding to the intersecting genomic regions from the healthy human hematopoiesis dataset. For this task, we used Signac's *CreateFragmentObject* and *FeatureMatrix* functions. Consequently, we performed the LSI on this set of peaks by employing Signac functions in the following order: *RunTFIDF* (with default parameters), *FindTopFeatures* (using 10% as the threshold for feature to be encompassed), and *RunSVD* (with n = *30*). Next, we used *FindTransferAnchors* function from the Seurat (v4.3.0) package[123] to find the anchor set between the reference and query dataset on shared peaks, using the following parameters: *reduction = 'lsiproject'*, *reference.reduction = 'lsi'*, and *dim = 2:30*, finding 3265 anchors. Finally, we ran *MapQuery* Seurat wrapper for transferring reference cell types onto the query cells, integrating both datasets, and projecting the query into the reference UMAP providing anchors and using *reference.reduction = 'lsi'* and *reduction.model = 'umap'* (loaded using *uwot* (v0.1.14)).

Trajectory analysis

To compute differentiation trajectories within the leukemic populations, we used ArchR's supervised trajectory tools. We first reasoned that three broadly different lineages were present within the leukemic populations. Next, we reasoned that the starting point for each differentiation trajectory should be the leukemic cluster with the highest LSC score and lowest CD38 gene accessibility, namely the LMPP-like cluster. This was further justified by our projection of healthy hematopoiesis reference. We generated each of the three trajectories with the *addTrajectory* function, which computes a pseudotime vector as described in the ArchR pipeline. To plot gene scores across trajectories, we used the *plotTrajectory* function with the *colorBy = 'GeneScoreMatrix'* option and *name = 'gene of interest'*. To plot clonal presence across pseudotime, the *geom_density* function from the ggplot2 (v3.4.0) package was used. To plot clone-specific gene activity across a common trajectory, *geom_smooth* with loess method was used.

We also computed the three trajectories in clone-specific manner, using genotype information, by retaining cells belonging only to one specific clone. We did this in order to understand how clone-specific dynamics of differentiation along the same lineage were reflected in chromatin accessibility of key TF motifs. To plot variation in motif accessibility of single motifs across trajectories, we used *plotTrajectory* with *colorBy = 'MotifMatrix'* and *name = 'motif of interest'*. To plot global TF motif deviation changes along these clone-specific trajectories, we used the *plotTrajectoryHeatmap* function, using the default smoothing window generating a smoothened pseudotime x motif matrix. This was used to generate clone-specific heatmaps of motif deviation change across pseudotime, which were then compared between different clones. To visualize how genomic elements changed accessibility across clone-specific trajectories, we split each clone-specific trajectory into five quantiles and used each quantile as pseudo-bulk for track plotting.

Motif correlation analysis

To correlate accessibility of TF motifs within the LMPP-like differentiation stage, we first extracted the per-cell z-scores computed by chromVAR from the peak matrix using *getMatrixFromProject*. We then subtracted these to contain only LMPP-like cells and plotted correlations of *Z* score for specific motifs. We used Spearman correlation to define whether there was a statistically significant positive or negative correlation.

Footprinting analysis

To perform footprinting analyses, we first added pseudo-bulk replicates via *addGroupCoverages*. We next obtained footprints for motifs of interest via *getFootprints* specifying the positions via *getPositions*. We used *plotFootprints* to generate the graphs. We substracted the Tn5 bias from the footprint to perform normalization.

Enrichment of genomic elements in differentially accessible peaks

To understand whether differentially accessible peaks of open chromatin were enriched in promoters versus distal elements or vice-versa, we joined differentially accessible peaks with peak metadata generated via *addReproduciblePeakSet.* We then filtered this list to retain only peaks mapping to promoter regions or to distal elements. Those peaks found in genomic regions ranging from 2kb upstream to 100bp downstream of annotated transcription start sites (default settings) were defined as promoter peaks. Peaks mapping to distal elements are all peaks that map neither to promoter regions nor to gene bodies (exons or introns). Fisher's exact test was used to estimate whether there was a significant overrepresentation of one peak type compared to the other, between sets of differentially accessible peaks.

Identification of putative PRC1.1 direct targets in AML from public data

We used previously published ChIP-seq data[77] to compile a list of putative direct PRC1.1 target genes in AML, defined as those genes bound by KDM2B (the DNA-binding component of PRC1.1.) in at least 5 out of 6 previously analyzed AML patients. We do note that ChIP-seq results may vary between patients, depending on cellular context and specific populations expanded (PRC1.1 binding may be dependent on cellular stage of differentiation); hence, this list of targets likely does not represent all possible targets, but rather those genes most frequently bound by PRC1.1 in AML.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details (type of statistical tests, exact sample size, dispersion and precision measures) are indicated in each figure legend. p values < *0.05* were considered statistically significant and are reported in the figures. Data were analyzed using GraphPad Prism 9 and R version 4.1.3. Detailed statistics related to Figures 4, 5, and 6 (and related Supplemental Figures) are reported in method details.

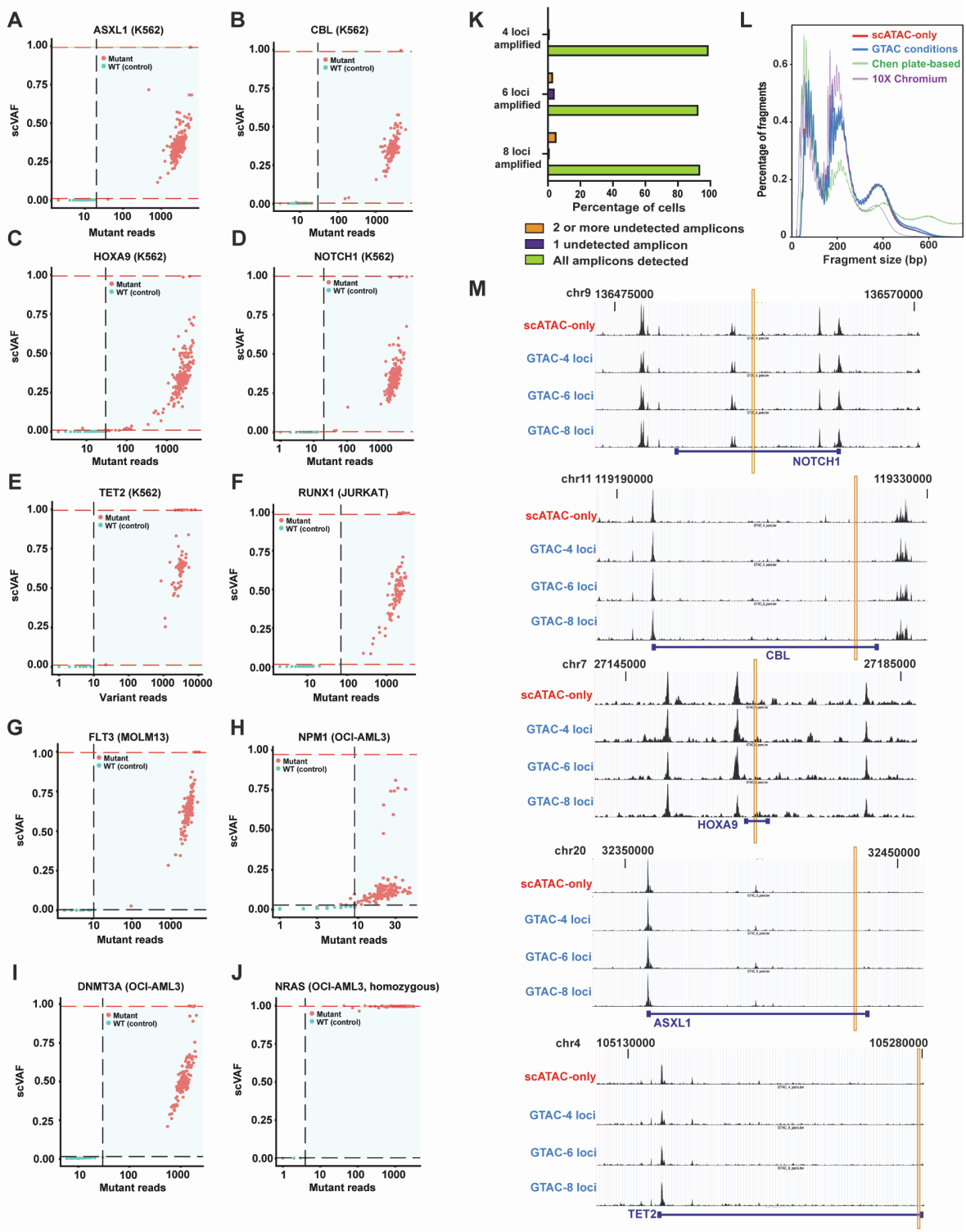
Cell Stem Cell, Volume *30*

Supplemental Information

GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells

Sven Turkalj, Niels Asger Jakobsen, Angus Groom, Marlen Metzner, Simone G. Riva, E. Ravza Gür, Batchimeg Usukhbayar, Mirian Angulo Salazar, Lance D. Hentges, Gerda Mickute, Kevin Clark, Paul Sopp, James O.J. Davies, Jim R. Hughes, and Paresh Vyas

Supplemental Figures and Legends.



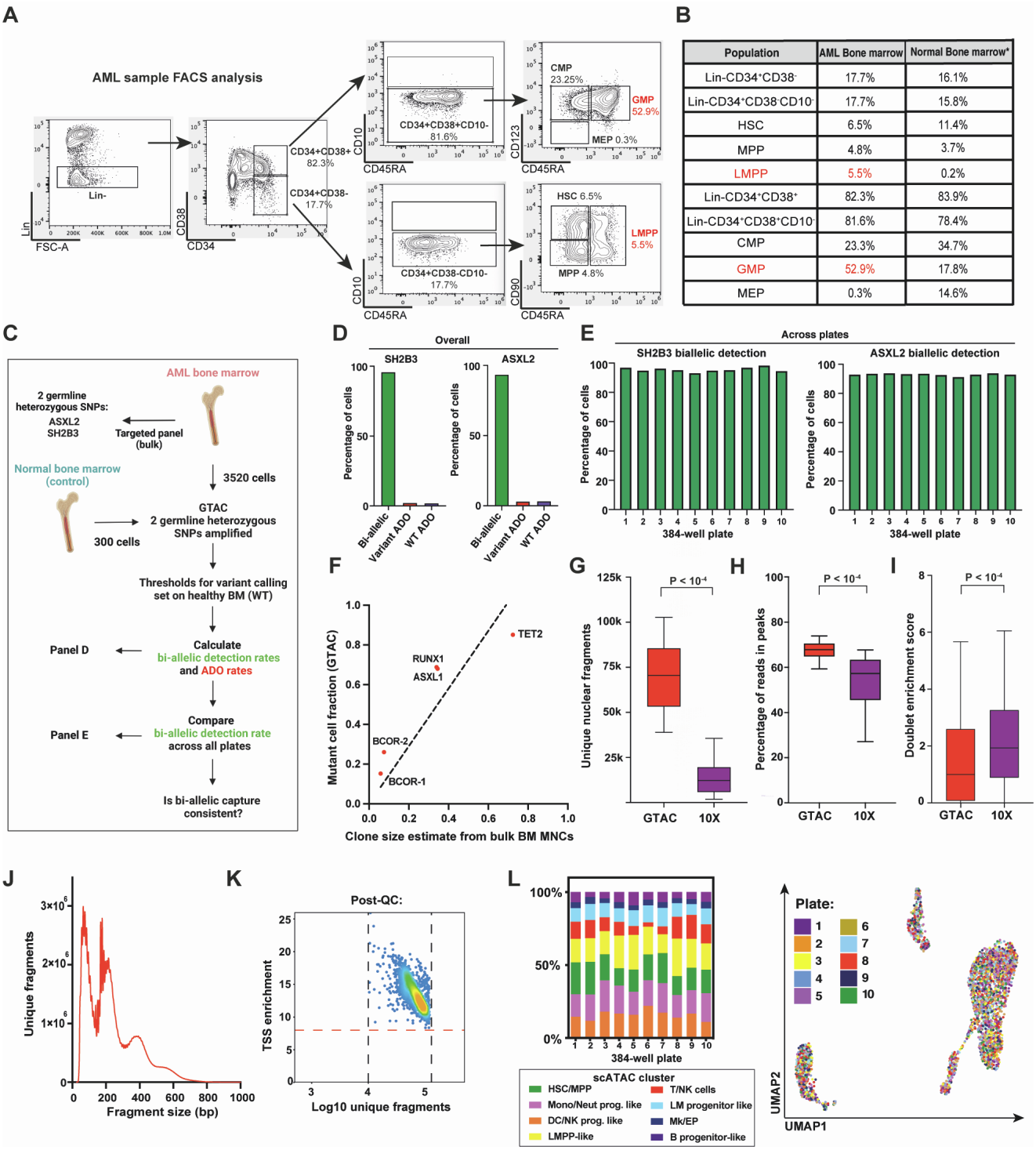
Supplemental Figure 1. Related to Figures 2 and 3. Thresholds used for allelic drop-out (ADO) estimation and coverage around genotyped loci in K562

(A)-(J) scVAF (y-axis) versus the number of mutant or variant genotyping reads (x-axis) for nine heterozygous mutant or variant loci (A)-(I) and one homozygous mutant locus (J). Each dot represents a single cell. THP1 cells were used as homozygous WT control for each locus (cyan). Mutant cells used (red) are indicated above each graph. In each panel, the vertical black dotted line indicates the minimum number of mutant or variant reads required for a cell to be called mutant or variant (STAR Methods). The lower horizontal red dotted line indicates the minimum scVAF required for a cell to be called mutant or variant (STAR Methods). The upper horizontal red dotted line indicates the maximum scVAF for which a cell is considered heterozygous mutant or variant (STAR Methods). For heterozygous loci, dots outside of the azure rectangle are cells in which we estimate that ADO occurred.

(K) Percentage of K562 cells (x-axis) in which amplified loci were detected when either 4 (top), 6 (middle) or 8 (bottom) loci were amplified. The legend below shows the number of multiple amplicons detected. For each amplicon, at least 30 reads were required for the locus to be considered detected.

(L) Percentage of ATAC fragments (y-axis) as a function of fragment size (x-axis) showing data for K562 assayed by GTAC, scATAC alone, published K562 data obtained from a plate-based scATAC-seq approach (Chen plate-based) and from a commercial droplet-based approach (10X Chromium).

(M) UCSC genome browser tracks displaying aggregated reads (y-axis) across five genes containing heterozygous mutations/variants targeted in K562 cells (top to bottom: *NOTCH1*, *CBL*, *HOXA9*, *ASXL1*, *TET2*), comparing the 4 conditions (labelled on the left). Low signal in coding regions confirms that exonic variant profiling from scATAC data alone is unfeasible. Yellow rectangles indicate the position of the heterozygous mutations or variants analysed in genotyping.



Population	AML Bone marrow	Normal Bone marrow*
Lin-CD34$^+$CD38$^-$	17.7%	16.1%
Lin-CD34$^+$CD38$^-$CD10$^-$	17.7%	15.8%
HSC	6.5%	11.4%
MPP	4.8%	3.7%
LMPP	5.5%	0.2%
Lin-CD34$^+$CD38$^+$	82.3%	83.9%
Lin-CD34$^+$CD38$^+$CD10$^-$	81.6%	78.4%
CMP	23.3%	34.7%
GMP	52.9%	17.8%
MEP	0.3%	14.6%

Supplemental Figure 2. Related to Figure 4. AML sample immunophenotype, ADO estimation and scATAC quality metrics.

(A) FACS analysis of the AML sample. Percentages of Lin-$CD34^+$ cells are indicated below populations. HSC: haematopoietic stem cell, MPP: multipotent progenitor, LMPP: lymphoid-primer multipotent progenitor, CMP: common myeloid progenitor, GMP: granulocyte-monocyte progenitor, MEP: megakaryocyte-erythroid progenitor.

(B) Table of % of immunophenotypic populations as a proportion of Lin^-CD34^+ in the AML sample (left) and normal bone marrow[115], showing expansion of the GMP and LMPP immunophenotypic compartments in the AML sample.

(C) Overview of ADO estimation in the AML sample. Bone marrow cells were isolated from a newly diagnosed male AML patient. Next generation sequencing was performed on leukemic bone marrow mononuclear cells using a panel of genes recurrently mutated in myeloid malignancy. 2 heterozygous germline SNPs in *ASXL2* and *SH2B3* genes were identified. These two loci were co-amplified in the GTAC experiment together with the 5 somatic mutant loci. The same 2 loci were also amplified in 300 homozygous WT CD34+ human bone marrow cells used as control for setting ADO thresholds (STAR Methods).

(D) Percentage of cells (y-axis) in which both variant and WT alleles were detected (green) compared to cells where only the WT (red; i.e. there was variant ADO) or variant (purple; i.e. there was WT ADO) allele was detected, for the *SH2B3* and *ASXL2* loci. Percentages of cells in which there was bi-allelic detection, variant ADO, or WT ADO, respectively, are: for *SH2B3*: 95.8%, 2.2%, and 2.0%; for *ASXL2*: 93.5%, 3.1%, and 3.4%.

(E) Percentage of cells in which there was bi-allelic detection of *SH2B3* (left) and *ASXL2* (right) loci across ten 384-well plates.

(F) Fraction of mutant cells as estimated by GTAC (y-axis) versus the expected clone size inferred from bulk sequencing of bone marrow mononuclear cells (x-axis). The dotted line is the linear regression curve. Each locus is indicated by a red dot.

(G) Number of unique nuclear scATAC fragments obtained per cell, comparing data from the AML sample processed with GTAC (red, n=3285) with data from five mixed phenotype acute leukemia (MPAL) samples processed with 10X Chromium[41] (purple, n=20059). Median values are: 70475 for GTAC; 12192 for 10X. Cells with >1000 and <120000 unique nuclear ATAC fragments were considered. *P* values obtained using two-tailed *t*-test.

(H) Percentage of ATAC reads in peaks, comparing data from the AML sample processed with GTAC (red, n=3285) with data from five mixed phenotype acute leukemia (MPAL) samples processed with 10X Chromium[41] (purple, n=20059). Median values are: 67.83% for

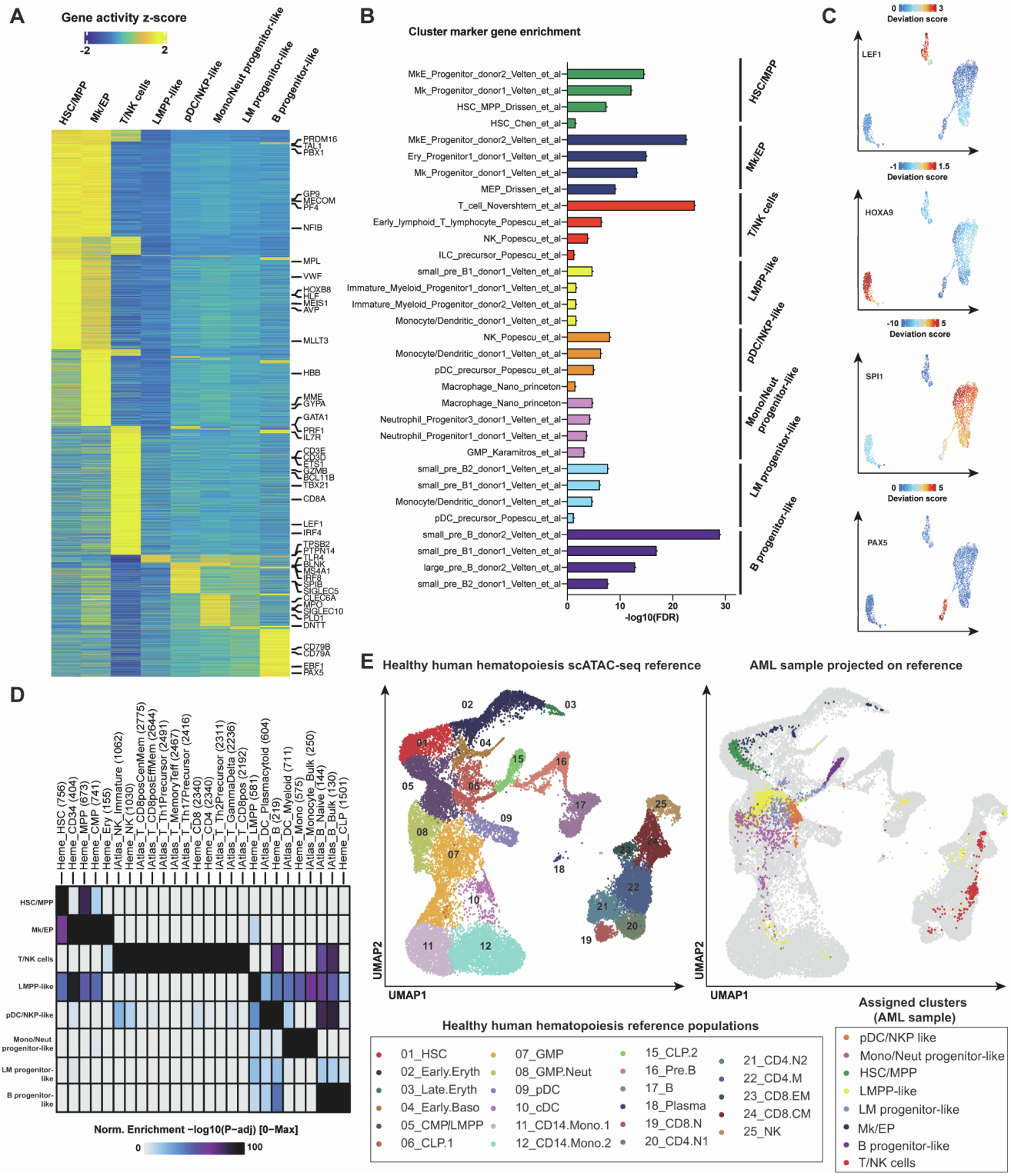
GTAC; 57.30% for 10X. Cells with >1000 and <120000 unique nuclear ATAC fragments were considered. *P* values obtained using two-tailed *t*-test.

(I) Doublet enrichment scores, comparing data from the AML sample processed with GTAC (red, n=3285) with data from five mixed phenotype acute leukemia (MPAL) samples processed with 10X Chromium[41] (purple, n=20059). Median values are: 1.00 for GTAC; 1.93 for 10X. Cells with >1000 and <120000 unique nuclear ATAC fragments were considered. *P* values obtained using two-tailed *t*-test.

(J) Number of ATAC fragments (y-axis) as a function of fragment size (x-axis) aggregated across all cells from our AML sample.

(K) Scatterplot showing the $\log_{10}$ of the number of unique nuclear fragments (x-axis) versus the TSS enrichment score (y-axis) for each cell. Cells with TSS enrichment < 8 (red horizontal dotted line), less than 10000 (left vertical dotted line) and more than 120000 (right vertical dotted line) unique reads were excluded. Cells passing QC are displayed as dots. High cell density is indicated in orange.

(L) Top, bars indicating the contribution of each plate (x-axis) to each cluster (colour coded). Below, UMAP from Figure 4G in which each cell is colour-coded by plate.


A
Gene activity z-score
-2
2
HSC/MPP
Mk/EP
T/NK cells
LMPP-like
pDC/NKP-like
Mono/Neut progenitor-like
LM progenitor-like
B progenitor-like
PRDM16
TAL1
PBX1
GP9
MECOM
PF4
NFIB
MPL
VWF
HOXB8
HLF
MEIS1
AVP
MLLT3
HBB
MME
GYPA
GATA1
PRF1
IL7R
CD3E
CD3D
ETS1
GZMB
BCL11B
TBX21
CD8A
LEF1
IRF4
TPSB2
PTPN14
TLR4
BLNK
MS4A1
IRF8
SPIB
SIGLEC5
CLEC6A
MPO
SIGLEC10
FLT3
DNTT
CD79B
CD79A
EBF1
PAX5
B
Cluster marker gene enrichment
MkE_Progenitor_donor2_Velten_et_al
Mk_Progenitor_donor1_Velten_et_al
HSC_MPP_Drissen_et_al
HSC_Chen_et_al
MkE_Progenitor_donor2_Velten_et_al
Ery_Progenitor1_donor1_Velten_et_al
Mk_Progenitor_donor1_Velten_et_al
MEP_Drissen_et_al
T_cell_Novershtern_et_al
Early_lymphoid_T_lymphocyte_Popescu_et_al
NK_Popescu_et_al
ILC_precursor_Popescu_et_al
small_pre_B1_donor1_Velten_et_al
Immature_Myeloid_Progenitor1_donor1_Velten_et_al
Immature_Myeloid_Progenitor_donor2_Velten_et_al
Monocyte/Dendritic_donor1_Velten_et_al
NK_Popescu_et_al
Monocyte/Dendritic_donor1_Velten_et_al
pDC_precursor_Popescu_et_al
Macrophage_Nano_princeton
Macrophage_Nano_princeton
Neutrophil_Progenitor3_donor1_Velten_et_al
Neutrophil_Progenitor1_donor1_Velten_et_al
GMP_Karamitros_et_al
small_pre_B2_donor1_Velten_et_al
small_pre_B1_donor1_Velten_et_al
Monocyte/Dendritic_donor1_Velten_et_al
pDC_precursor_Popescu_et_al
small_pre_B_donor2_Velten_et_al
small_pre_B1_donor1_Velten_et_al
large_pre_B_donor2_Velten_et_al
small_pre_B2_donor1_Velten_et_al
HSC/MPP
Mk/EP
T/NK cells
LMPP-like
pDC/NKP-like
Mono/Neut progenitor-like
LM progenitor-like
B progenitor-like
0
10
20
30
-log10(FDR)
C
0
3
Deviation score
LEF1
-1
1.5
Deviation score
HOXA9
-10
5
Deviation score
SPI1
0
5
Deviation score
PAX5
D
Heme_HSC (756)
Heme_CD34 (404)
Heme_MPP (673)
Heme_CMP (741)
Heme_Ery (155)
IAtlas_NK_Immature (1062)
Heme_NK (1030)
IAtlas_T_CD8posCenMem (2775)
IAtlas_T_CD8posEffMem (2644)
IAtlas_T_Th1Precursor (2491)
IAtlas_T_MemoryTeff (2467)
IAtlas_T_Th17Precursor (2416)
Heme_CD8 (2340)
Heme_CD4 (2340)
IAtlas_T_Th2Precursor (2311)
IAtlas_T_GammaDelta (2236)
IAtlas_T_CD8pos (2192)
Heme_LMPP (581)
IAtlas_DC_Plasmacytoid (604)
Heme_B (219)
IAtlas_DC_Myeloid (711)
Heme_Mono (575)
IAtlas_Monocyte_Bulk (250)
IAtlas_B_Naive (144)
IAtlas_B_Bulk (130)
Heme_CLP (1501)
HSC/MPP
Mk/EP
T/NK cells
LMPP-like
pDC/NKP-like
Mono/Neut progenitor-like
LM progenitor-like
B progenitor-like
Norm. Enrichment −log10(P−adj) [0−Max]
0
100
E
Healthy human hematopoiesis scATAC-seq reference
AML sample projected on reference
UMAP1
UMAP2
Healthy human hematopoiesis reference populations
01_HSC
02_Early.Eryth
03_Late.Eryth
04_Early.Baso
05_CMP/LMPP
06_CLP.1
07_GMP
08_GMP.Neut
09_pDC
10_cDC
11_CD14.Mono.1
12_CD14.Mono.2
15_CLP.2
16_Pre.B
17_B
18_Plasma
19_CD8.N
20_CD4.N1
21_CD4.N2
22_CD4.M
23_CD8.EM
24_CD8.CM
25_NK
Assigned clusters (AML sample)
pDC/NKP like
Mono/Neut progenitor-like
HSC/MPP
LMPP-like
LM progenitor-like
Mk/EP
B progenitor-like
T/NK cells

Supplemental Figure 3. Related to Figure 4. Cluster annotation in GTAC data from the primary AML sample.

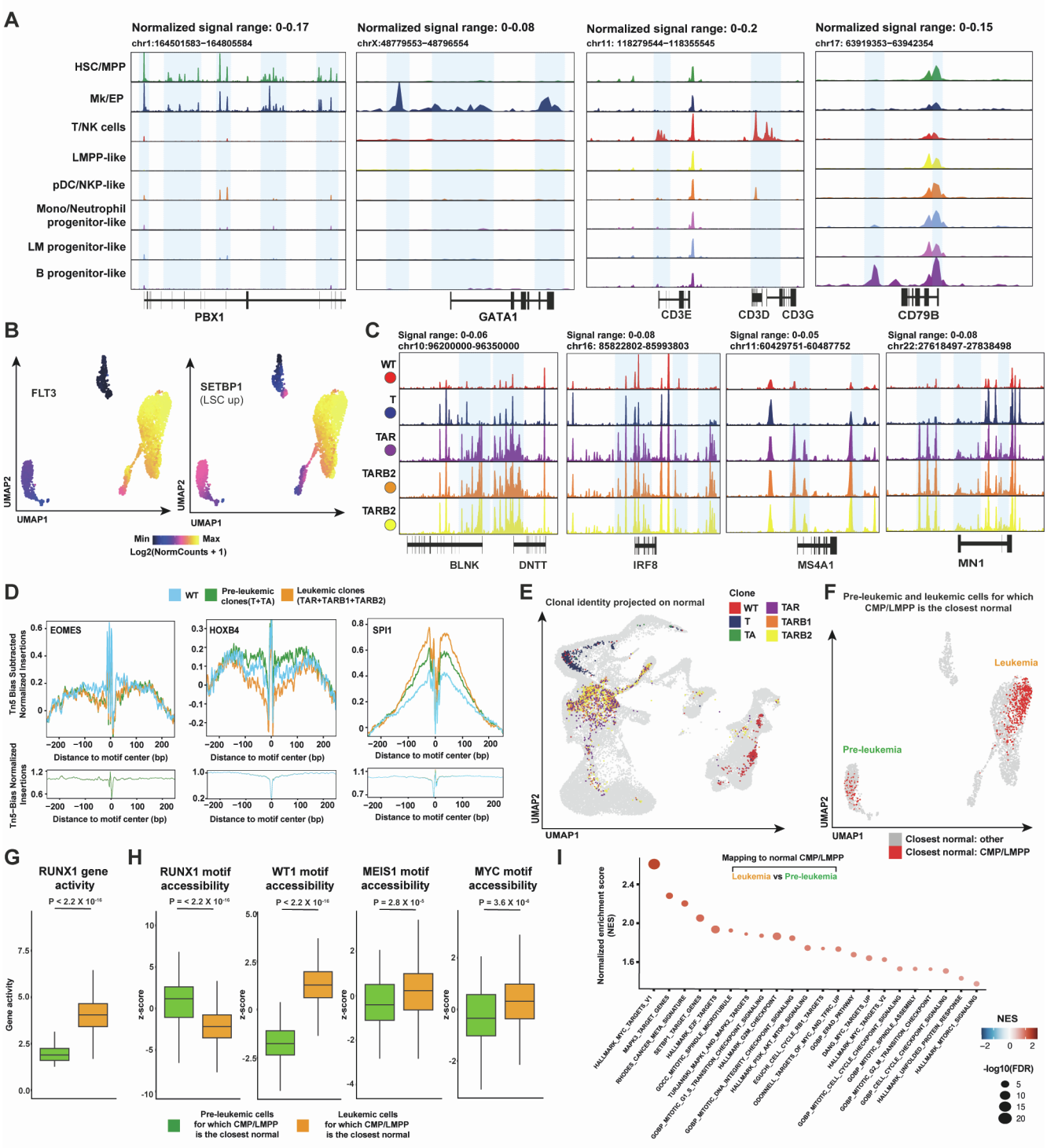
(A) Heatmap of gene activity scores for marker genes identified for each cluster (columns). Each row represents a gene. Marker genes with FDR < 0.05 and $log_2FC > 0.5$ are shown. Several key marker genes are labelled on the right.

(B) Enrichment of haematopoietic gene signatures within marker genes identified for each cluster, based on gene activity scores. Here, marker genes were selected with FDR < 0.05 and $log_2FC > 0.25$. Clusters are labelled on the right. Bars are colour-coded according to cluster, as in Figure 4G. Signatures were compiled using published scRNA-seq data[113-118]. For each cluster, the top 4 enriched signatures are shown.

(C) Top to bottom: the UMAP from Figure 4G overlaid with per-cell deviation scores for motif accessibility for the LEF1, HOXA9, SPI1, and PAX5 motifs.

(D) Heatmap showing enrichment of curated peak sets from previously published bulk ATAC-seq data (provided by ArchR) within marker peaks identified for each cluster. Marker peaks were selected with FDR < 0.05 and $log_2FC > 0.25$. The colour scale indicates the normalized peak set enrichment ($-log_{10}$(adjusted *P* value)) on a scale of 0-100. The maximum $-log_{10}$(adjusted *P* value) for each peak set is indicated by the numbers in brackets following the name of the enriched ATAC annotation.

(E) Left: UMAP based on previously published[41] scATAC-seq data from healthy human hematopoiesis. Cluster labelling is indicated below the graphs. Numbers on the graph indicate the healthy reference bone marrow population. Clusters and labels were kept identical as in the original publication. Right: Projection of single cells from our AML dataset, coloured by assigned cluster, onto the healthy human hematopoiesis reference (grey).


A
Normalized signal range: 0-0.17
chr1:164501583−164805584
Normalized signal range: 0-0.08
chrX:48779553−48796554
Normalized signal range: 0-0.2
chr11: 118279544−118355545
Normalized signal range: 0-0.15
chr17: 63919353−63942354
HSC/MPP
Mk/EP
T/NK cells
LMPP-like
pDC/NKP-like
Mono/Neutrophil progenitor-like
LM progenitor-like
B progenitor-like
PBX1
GATA1
CD3E
CD3D
CD3G
CD79B
B
FLT3
SETBP1
(LSC up)
UMAP2
UMAP1
Min
Max
Log2(NormCounts + 1)
C
Signal range: 0-0.06
chr10:96200000-96350000
Signal range: 0-0.08
chr16: 85822802-85993803
Signal range: 0-0.05
chr11:60429751-60487752
Signal range: 0-0.08
chr22:27618497-27838498
WT
T
TAR
TARB2
TARB2
BLNK
DNTT
IRF8
MS4A1
MN1
D
WT
Pre-leukemic clones(T+TA)
Leukemic clones (TAR+TARB1+TARB2)
EOMES
HOXB4
SPI1
Tn5 Bias Subtracted Normalized Insertions
Tn5−Bias Normalized Insertions
Distance to motif center (bp)
E
Clonal identity projected on normal
Clone
WT
T
TA
TAR
TARB1
TARB2
UMAP2
UMAP1
F
Pre-leukemia and leukemic cells for which CMP/LMPP is the closest normal
Leukemia
Pre-leukemia
UMAP2
UMAP1
Closest normal: other
Closest normal: CMP/LMPP
G
RUNX1 gene activity
P < 2.2 X 10-16
Gene activity
H
RUNX1 motif accessibility
P = < 2.2 X 10-16
WT1 motif accessibility
P < 2.2 X 10-16
MEIS1 motif accessibility
P = 2.8 X 10-3
MYC motif accessibility
P = 3.6 X 10-6
z-score
Pre-leukemic cells for which CMP/LMPP is the closest normal
Leukemic cells for which CMP/LMPP is the closest normal
I
Mapping to normal CMP/LMPP
Leukemia vs Pre-leukemia
Normalized enrichment score (NES)
HALLMARK_MYC_TARGETS_V1
MAPK3_TARGET_GENES
RHODES_CANCER_META_SIGNATURE
SETDP1_TARGET_GENES
HALLMARK_E2F_TARGETS
GOCC_MITOTIC_SPINDLE_MICROTUBULE
TURJANSKI_MAPK1_AND_MAPK2_TARGETS
GOBP_MITOTIC_G1_S_TRANSITION_CHECKPOINT_SIGNALING
GOBP_MITOTIC_DNA_INTEGRITY_CHECKPOINT_SIGNALING
HALLMARK_G2M_CHECKPOINT
HALLMARK_PI3K_AKT_MTOR_SIGNALING
EGUCHI_CELL_CYCLE_RB1_TARGETS
ODONNELL_TARGETS_OF_MYC_AND_TFRC_UP
GOBP_ERAD_PATHWAY
DANG_MYC_TARGETS_UP
HALLMARK_MYC_TARGETS_V1
GOBP_MITOTIC_CELL_CYCLE_CHECKPOINT_SIGNALING
GOBP_MITOTIC_SPINDLE_ASSEMBLY
GOBP_MITOTIC_G2_M_TRANSITION_CHECKPOINT
GOBP_CELL_CYCLE_CHECKPOINT_SIGNALING
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_MITOTIC_SPINDLE
NES
-log10(FDR)

Supplemental Figure 4. Related to Figure 4. A homozygous *RUNX1* mutation is associated with progression from preleukemia to leukemia and correlates with profound differences in chromatin states.

(A) Normalized scATAC sequencing read coverage (y-axis), comparing cluster pseudo-bulks (labelled on the left) around the *PBX1*, *GATA1, CD3E/D/G*, and *CD79B* loci. Signal range is displayed above each graph. Horizontal tracks represent cluster pseudo-bulks. Chromosome coordinates are indicated above. Gene annotations are indicated below.

(B) The UMAP from Figure 4G overlaid with gene activity scores for *FLT3* and *SETBP1*.

(C) Normalized scATAC sequencing read coverage (y-axis), comparing clone pseudo-bulks around the *BLNK/DNTT, IRF8*, *MS4A1* and *MN1* loci. Each horizontal track represents a different clone (labelled on the left). Rest of the panel as in (A).

(D) TF footprints for the EOMES, HOXB4 and SPI1 motifs comparing WT (azure, n=377 cells), preleukemic (green, n=595 cells) and leukemic clones (orange, n=2141 cells). Top, Tn5 bias-subtracted signal (y-axis) as a function of distance from the motif centre (x-axis). Bottom, Tn5 binding bias (y-axis) as a function of distance from the motif centre (x-axis), based on bias prediction computed by screening coverage of random hexamers within Tn5 insertion sites, using pseudo-bulk replicates. 1 indicates no bias. Colour legend is indicated above.

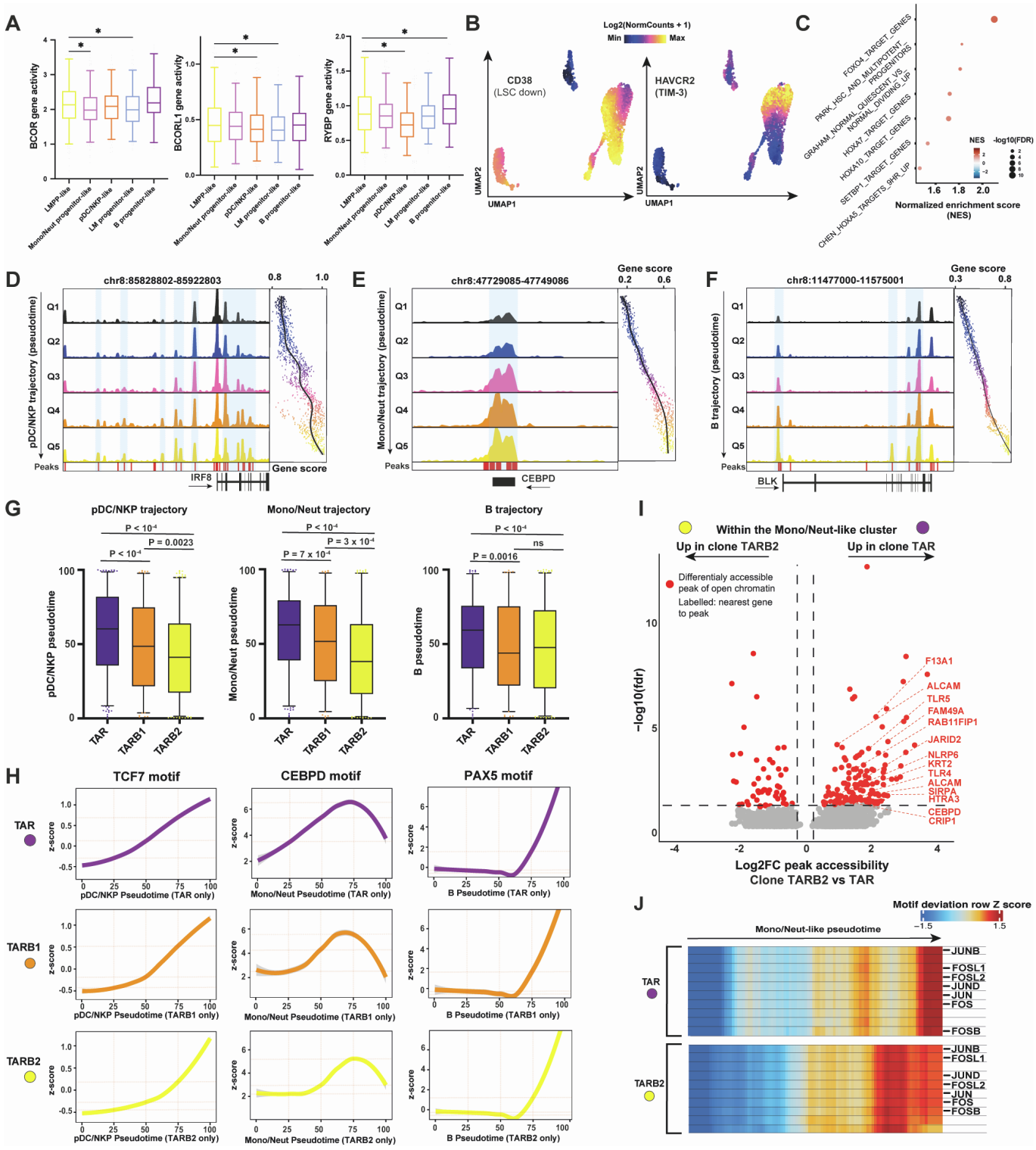
(E) Projection of single cells from our AML dataset, coloured by clone, onto the previously published[41] scATAC-seq healthy human hematopoiesis reference (grey, same as in Figure S3E). Colour legend for clones is labelled in the top right corner (same as in Figure 4).

(F) UMAP from Figure 4G, highlighting cells that mapped onto normal CMP/LMPP in red (all other cells are in grey).

(G) Boxplots comparing *RUNX1* gene activity between preleukemic (green, n=97) and leukemic (orange, n=492) cells for which the projected closest normal population was CMP/LMPP. Wilcoxon rank sum test was used. The *P* value is reported above.

(H) Same as (G) but comparing RUNX1, WT1, MEIS1 and MYC binding motif accessibility between the two groups.

(I) Gene set enrichment analysis (GSEA) of genes with increased activity in leukemic cells compared to preleukemic cells within cells for which the closest normal population was CMP/LMPP. The y-axis plots the normalized enrichment score (NES). Dot size indicates $-\log_{10}$(FDR). Key pre-selected enriched pathways are indicated below (x-axis).


A
BCOR gene activity
BCORL1 gene activity
RYBP gene activity
LMPP-like
Mono/Neut progenitor-like
pDC/NKP-like
LM progenitor-like
B progenitor-like
B
Log2(NormCounts + 1)
Min
Max
CD38
(LSC down)
HAVCR2
(TIM-3)
UMAP1
UMAP2
C
FOXO4_TARGET_GENES
PARK_HSC_AND_MULTIPOTENT_PROGENITORS
GRAHAM_NORMAL_QUIESCENT_VS_NORMAL_DIVIDING_UP
HOXA7_TARGET_GENES
HOXA10_TARGET_GENES
SETBP1_TARGET_GENES
CHEN_HOXA5_TARGETS_9HR_UP
NES
-log10(FDR)
Normalized enrichment score (NES)
D
chr8:85828802-85922803
pDC/NKP trajectory (pseudotime)
Q1
Q2
Q3
Q4
Q5
Peaks
Gene score
IRF8
E
chr8:47729085-47749086
Mono/Neut trajectory (pseudotime)
CEBPD
F
chr8:11477000-11575001
B trajectory (pseudotime)
BLK
G
pDC/NKP trajectory
Mono/Neut trajectory
B trajectory
P < 10-4
P = 0.0023
P = 7 x 10-4
P = 3 x 10-4
P = 0.0016
ns
pDC/NKP pseudotime
Mono/Neut pseudotime
B pseudotime
TAR
TARB1
TARB2
H
TCF7 motif
CEBPD motif
PAX5 motif
z-score
pDC/NKP Pseudotime (TAR only)
Mono/Neut Pseudotime (TAR only)
B Pseudotime (TAR only)
pDC/NKP Pseudotime (TARB1 only)
Mono/Neut Pseudotime (TARB1 only)
B Pseudotime (TARB1 only)
pDC/NKP Pseudotime (TARB2 only)
Mono/Neut Pseudotime (TARB2 only)
B Pseudotime (TARB2 only)
I
Within the Mono/Neut-like cluster
Up in clone TARB2
Up in clone TAR
Differentialy accessible peak of open chromatin
Labelled: nearest gene to peak
−log10(fdr)
F13A1
ALCAM
TLR5
FAM49A
RAB11FIP1
JARID2
NLRP6
KRT2
TLR4
ALCAM
SIRPA
HTRA3
CEBPD
CRIP1
Log2FC peak accessibility
Clone TARB2 vs TAR
J
Motif deviation row Z score
Mono/Neut-like pseudotime
JUNB
FOSL1
FOSL2
JUND
JUN
FOS
FOSB

Supplemental Figure 5. Related to Figure 5. GTAC identifies clone-specific chromatin dynamics throughout the leukemic hierarchy, highlighting a differentiation delay of *BCOR* mutant clones at an LSC-like state.

(A) Gene activity scores of genes encoding three PRC1.1 members (y-axis), namely *BCOR*, *BCORL1* and *RYBP*, across 5 leukemic clusters (labelled below). *P* values obtained using Kruskall-Wallis test with Dunnett's multiple comparisons. Significant differences are indicated with a star.

(B) The UMAP from Figure 4G overlaid with gene activity scores for *CD38* and *HAVCR2* (encoding the TIM-3 marker).

(C) Dot plot showing results of gene set enrichment analysis (GSEA), when LMPP-like cells were compared to all other leukemic cells. The x-axis plots the normalized enrichment score. Dot size indicates $-\log_{10}$(FDR). Pathways are indicated on the left (y-axis).

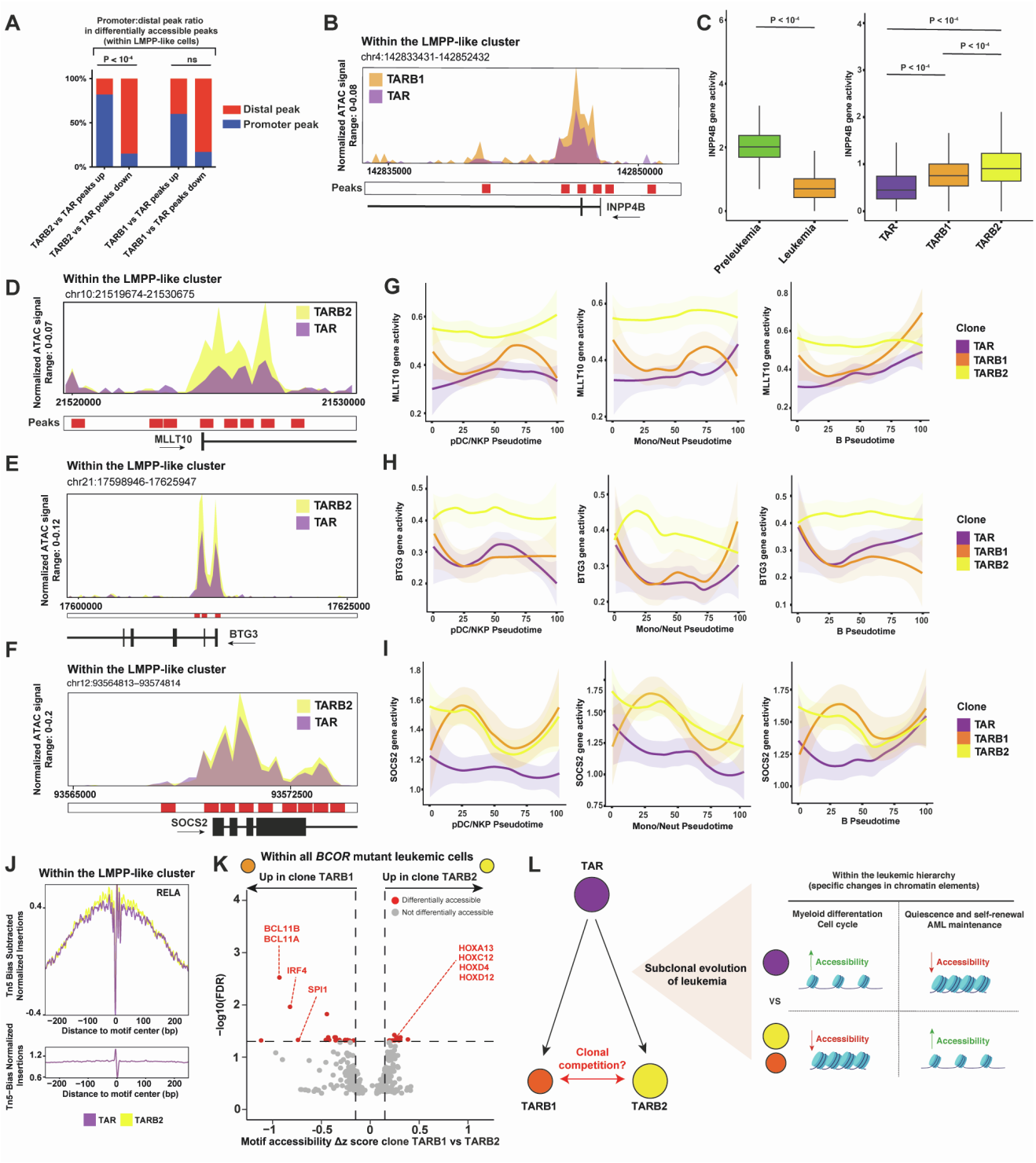
(D)-(F) Normalized scATAC sequencing read coverage around *IRF8* (D), *CEBPE* (E) and *BLK* (F) loci, across the pDC/NKP, Mono/Neut and B trajectory, respectively. Each trajectory was split into five quantiles (labelled Q1-Q5, colour-coded). Gene score as function of pseudotime is plotted on the right of each graph. Peaks and gene annotations are plotted below. Highlighted areas show variation across pseudotime.

(G) Boxplots depicting the per-cell pseudotime score (y-axis) across the three leukemic clones (colour-coded), for the three trajectories (left-to-right, indicated above each graph). *P* values obtained using Wilcoxon rank sum test.

(**H**) Same as in Figure 5E but showing TCF7, CEBPD and PAX5 motif accessibility.

(I) Volcano plot illustrating differentially accessible genomic peaks (FDR < 0.05 and $\log_2$FC > 0.25; labelled as red dots) between TARB2 (n=293 cells) and TAR (n=167 cells) clones within the Mono/Neut progenitor-like cluster. Vertical dotted lines indicate $\log_2$FC thresholds (-0.25 and 0.25), the horizontal dotted line indicates FDR=0.05. Labels indicate the closest gene to the peak. Labelled genes belong to monocytic and neutrophilic progenitor signatures.

(**J**) Heatmaps showing the variation of motif accessibility across clone-specific Mono/Neut pseudotime (x-axis), for a selected subset of most variable motifs across pseudotime (each row represents a motif). The upper heatmap relates to the TAR clone and the lower heatmap to the TARB2 clone. Motifs related to the AP-1 complex are labelled.


A
Promoter:distal peak ratio
in differentially accessible peaks
(within LMPP-like cells)
P < 10⁻⁴
ns
100%
50%
0%
Distal peak
Promoter peak
TARB2 vs TAR peaks up
TARB2 vs TAR peaks down
TARB1 vs TAR peaks up
TARB1 vs TAR peaks down
B
Within the LMPP-like cluster
chr4:142833431-142852432
Normalized ATAC signal
Range: 0-0.08
TARB1
TAR
142835000
142850000
Peaks
INPP4B
C
P < 10⁻⁴
INPP4B gene activity
Preleukemia
Leukemia
TAR
TARB1
TARB2
D
Within the LMPP-like cluster
chr10:21519674-21530675
Normalized ATAC signal
Range: 0-0.07
TARB2
TAR
21520000
21530000
Peaks
MLLT10
E
Within the LMPP-like cluster
chr21:17598946-17625947
Normalized ATAC signal
Range: 0-0.12
TARB2
TAR
17600000
17625000
BTG3
F
Within the LMPP-like cluster
chr12:93564813-93574814
Normalized ATAC signal
Range: 0-0.2
TARB2
TAR
93565000
93572500
SOCS2
G
MLLT10 gene activity
pDC/NKP Pseudotime
Mono/Neut Pseudotime
B Pseudotime
Clone
TAR
TARB1
TARB2
H
BTG3 gene activity
pDC/NKP Pseudotime
Mono/Neut Pseudotime
B Pseudotime
Clone
TAR
TARB1
TARB2
I
SOCS2 gene activity
pDC/NKP Pseudotime
Mono/Neut Pseudotime
B Pseudotime
Clone
TAR
TARB1
TARB2
J
Within the LMPP-like cluster
RELA
Tn5 Bias Subtracted Normalized Insertions
Distance to motif center (bp)
Tn5−Bias Normalized Insertions
Distance to motif center (bp)
TAR
TARB2
K
Within all BCOR mutant leukemic cells
Up in clone TARB1
Up in clone TARB2
Differentially accessible
Not differentially accessible
BCL11B
BCL11A
IRF4
SPI1
HOXA13
HOXC12
HOXD4
HOXD12
−log10(FDR)
Motif accessibility Δz score clone TARB1 vs TARB2
L
TAR
Clonal competition?
TARB1
TARB2
Subclonal evolution of leukemia
Within the leukemic hierarchy
(specific changes in chromatin elements)
Myeloid differentation
Cell cycle
Quiescence and self-renewal
AML maintenance
Accessibility
VS

Supplemental Figure 6. Related to Figure 6. GTAC reveals specific switches in chromatin accessibility within the LMPP-like differentiation stage upon subclonal acquisition of *BCOR* mutations.

(A) Ratio of differentially accessible peaks between *BCOR* mutant and *BCOR* WT LMPP-like cells which map to either distal elements (red) or promoters (blue). Two bars on the left represent the TARB2 versus TAR comparison. Two bars on the right represent the TARB1 versus TAR comparison. In each set of two bars, the left bar indicates peaks more accessible in *BCOR* mutant cells, and the right bar indicates peaks more accessible in *BCOR* WT cells. Fisher's exact test was performed to determine significance (ns: non significant).

(B) Same as Figure 6E but comparing the TAR and TARB1 clones.

(C) Left: Boxplots comparing *INPP4B* gene activity between genetic preleukemic and genetic leukemic cells. Wilcoxon rank sum test was used. The *P* value is reported above. Right: Boxplots comparing *INPP4B* gene activity between TAR, TARB1 and TARB2 leukemic clones. *P* values obtained using Kruskal-Wallis test with Dunn's test for multiple comparisons.

(**D**)-(**F)** Same as in Figure 5E but around the *MLLT10* (C)*, BTG3* (D), and *SOCS2* (E) gene loci, focusing on TSS-proximal regions.

(**G**)-(**I**) Same as in Figure 5F but relative to *MLLT10* (F), *BTG3* (G), and *SOCS2* (H) gene activity.

(J) TF footprint for the RELA motif comparing TARB2 (yellow) and TAR (purple) LMPP-like cells. Top, Tn5 bias-subtracted signal (y-axis) as a function of distance from the motif centre (x-axis). Bottom, Tn5 binding bias (y-axis) as a function of distance from the motif centre (x-axis), based on bias prediction computed by screening coverage of random hexamers within Tn5 insertion sites, using pseudo-bulk replicates. 1 indicates no bias.

(K) Volcano plot of differential TF motif accessibility between TARB1 (left) and TARB2 (right) clones across all leukemic clusters. Differentially accessible motifs (FDR < 0.05 and absolute $\Delta z > 0.20$) are labelled as red dots. Vertical dotted lines indicate Δz thresholds, the horizontal dotted line indicates the FDR threshold.

(L) Graphical depiction of the hypothesis regarding clonal dominance of *BCOR* mutant clones within the leukemic hierarchy, based on GTAC data from this patient.

A.4 A protocol for simultaneous high-sensitivity genotyping and chromatin accessibility profiling in single cells

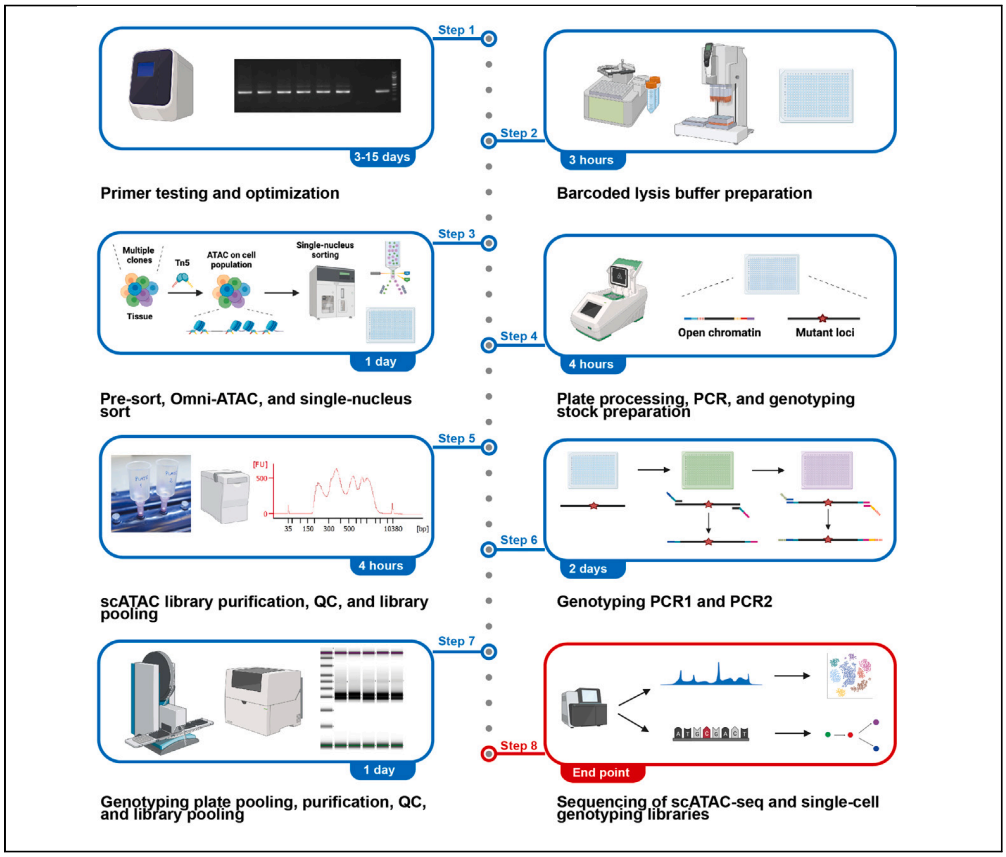
pdfpages

STAR Protocols

CellPress
OPEN ACCESS

Protocol

A protocol for simultaneous high-sensitivity genotyping and chromatin accessibility profiling in single cells



Single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) resolves the heterogeneity of epigenetic states across cells but does not typically capture exonic mutations, which limits our knowledge of how somatic mutations alter chromatin landscapes. Here, we present a plate-based approach coupling high-sensitivity genotyping of genomic loci with high-content scATAC-seq libraries from the same single cells. We first describe steps for optimization of genotyping primers, followed by detailed guidance on the preparation of both scATAC-seq and single-cell genotyping libraries, fully automated on high-throughput liquid handling platforms.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Sven Turkalj, Niels Asger Jakobsen, Angus Groom, Felix A. Radtke, Paresh Vyas

sven.turkalj@rdm.ox.ac.uk (S.T.)
paresh.vyas@imm.ox.ac.uk (P.V.)

Highlights

scATAC-seq coupled with high-sensitivity mutation capture at multiple genomic loci

Link genetic and epigenetic evolution in malignant and pre-malignant tissues

Guidance on barcoding, automation, and liquid handling for high-throughput applications

Turkalj et al., STAR Protocols 4, 102641
December 15, 2023 Crown Copyright © 2023
https://doi.org/10.1016/j.xpro.2023.102641

STAR Protocols

CellPress
OPEN ACCESS

Protocol

A protocol for simultaneous high-sensitivity genotyping and chromatin accessibility profiling in single cells

Sven Turkalj,[1,2,4,6,*] Niels Asger Jakobsen,[1,2,3,4] Angus Groom,[1,2,5] Felix A. Radtke,[1,2,5] and Paresh Vyas[1,2,3,7,*]

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

[2]Oxford Centre for Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[3]Department of Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[4]These authors contributed equally

[5]These authors contributed equally

[6]Technical contact

[7]Lead contact

*Correspondence: sven.turkalj@rdm.ox.ac.uk (S.T.), paresh.vyas@imm.ox.ac.uk (P.V.)
https://doi.org/10.1016/j.xpro.2023.102641

SUMMARY

Single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) resolves the heterogeneity of epigenetic states across cells but does not typically capture exonic mutations, which limits our knowledge of how somatic mutations alter chromatin landscapes. Here, we present a plate-based approach coupling high-sensitivity genotyping of genomic loci with high-content scATAC-seq libraries from the same single cells. We first describe steps for optimization of genotyping primers, followed by detailed guidance on the preparation of both scATAC-seq and single-cell genotyping libraries, fully automated on high-throughput liquid handling platforms.

For complete details on the use and execution of this protocol, please refer to Turkalj, Jakobsen et al.[1]

BEFORE YOU BEGIN

Institutional permissions

Patient samples were collected with informed consent under ethically approved protocols (MREC 06/Q1606/110 or NHS REC 17/YH/0382). Written informed consent was obtained in accordance with the Declaration of Helsinki. If GTAC is performed on primary human samples, before the collection of clinical biopsies, the study requires the approval by the Ethics Committee or by an equivalent organization. Informed consent should be collected from all patients.

Preparation one: Determination of the optimal number of PCR cycles for the specific tissue type

Timing: 2 days

Before the execution of GTAC on primary samples of interest, we strongly recommend performing a pilot experiment on a sample derived from the same tissue type first, to define the optimal number of PCR cycles required for single-cell amplification of open chromatin fragments (part 3, point 36). For preparation one, it is sufficient to perform plate-based scATAC-seq,[2,4] without the addition of target-specific genotyping primers. During these steps, you will sort single tagmented nuclei into multiple 96-well plates. One plate will be used for preparation one, to define the optimal number

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

of PCR cycles. The remaining plates (stored at −80°C) will be used in preparation three, to test genotyping primers in single cells.

⚠ CRITICAL: These pilot experiments can help avoid several potential issues (troubleshooting).

Note: PCR cycle numbers stated throughout the protocol apply to healthy and leukemic primary human bone marrow mononuclear cells, but these numbers may vary across tissue types.

Note: Before starting with preparation one, prepare all buffers listed in materials and equipment. Order generic i7 and i5 primers and reconstitute them at 100 μM in TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA).

Optional: Depending on the tissue type (solid versus cell suspension) and preservation state (fresh versus cryopreserved) of your sample, you may need to prepare different buffers for sample preparation. Follow the considerations in part 2, point 7 and prepare buffers according to the sample type. If FACS staining and pre-sorting prior to Omni-ATAC are required, prepare FACS buffer, thawing media, and antibody and fluorescence-minus one (FMO) mixes.

Note: Points 1, 4, and 5a should be performed in a designated pre-PCR clean area, ideally in a biosafety cabinet, prior to PCR.

1. Prepare 2 mL of generic i7 and i5 20 μM primer dilutions in nuclease-free water. Prepare generic lysis buffer for 10 96-well plates, according to the table below.

Generic lysis buffer		
Reagent	1 reaction	1150 reactions (10 96-well plates + 20% dead volume)
Tris HCl pH 8.0 1 M	0.2 μL	230 μL
NaCl 5 M	0.04 μL	46 μL
SDS 10% (wt/vol)	0.08 μL	92 μL
Generic i7 primer 20 μM	1 μL	1150 μL
Generic i5 primer 20 μM	1 μL	1150 μL
Nuclease-free water	1.68 μL	1932 μL
TOTAL	**4 μL**	**4600 μL**

   a. Aliquot 4 μL of generic lysis buffer into each well of a fresh 96-well plate.
   b. When a plate is aliquoted, cover it with aluminum adhesive seals, centrifuge in a plate spinner for 10 s at 500 × *g*, and place on wet ice. Repeat for all plates.
2. Prepare the sample and perform Omni-ATAC as detailed in part 2, points 7–26.

   Note: Follow the considerations in part 2, point 7, regarding the optimal course of action for different tissue types!

3. Sort single tagmented nuclei into plates prepared in point 1:
   a. For sorter calibration and sorting panel setup, follow part 2, points 27–29, but calibrate the machine for sorting into 96-well plates.
   b. Centrifuge lysis buffer plates in a plate spinner for 10 s at 500 × *g*.
   c. Sort single nuclei in each column, keeping 1–2 wells of the plate empty.
   d. After nuclei have been sorted into the whole plate, cover it with an aluminum adhesive seal, centrifuge immediately in a plate spinner for 20 s at 500 × *g*, and snap freeze the plate by placing it on dry ice.

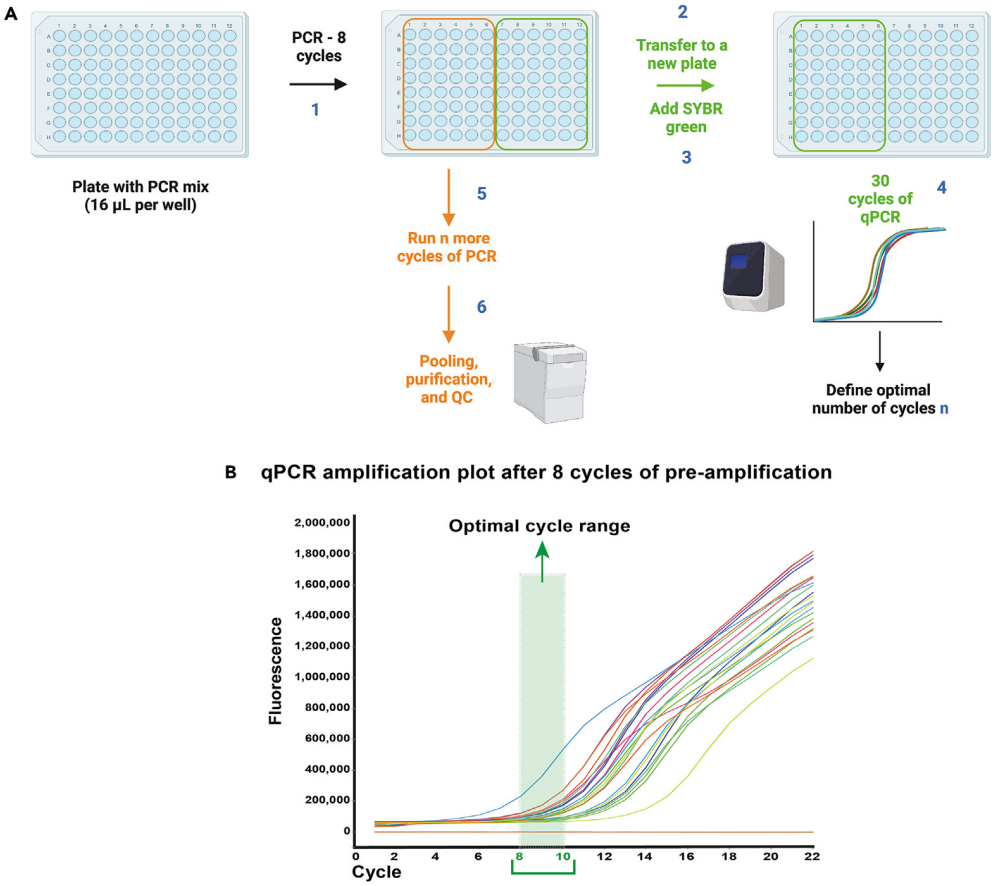


Figure 1. Optimal PCR cycle number titration

(A) Procedures for establishing the optimal number of PCR cycles for a GTAC experiment. Blue numbers indicate the order in which the steps are executed.

(B) SYBR Green-based qPCR was performed after 8 initial cycles of amplification. Each curve represents a single well. The optimal additional number of PCR cycles is highlighted in green.

Pause point: After the sort, plates can be stored at −80°C for up to 3 months. Some sorted plates will also be used in preparation three.

4. Using one sorted 96-well plate, proceed with lysis and Tween 20 quenching:
   a. Incubate the plate at 65°C for 15 min in a thermal cycler, with the lid heated to 105°C.
   b. Add 2 µL of Tween 20 10% (vol/vol) to each well using an 8 × 20 µL multichannel pipette.
   c. Centrifuge the plate in a plate spinner for 10 s at 500 × *g* and incubate at 21°C for 10 min.
5. Prepare the generic PCR mix by mixing 960 µL of Q5 High-Fidelity 2× Master Mix with 240 µL of nuclease-free water.
   a. Dispense 10 µL of this mix to each well using an 8 × 20 µL multichannel pipette. Dispense on the sides of the wells and do not touch the bottom with the tips. Keep the plate and the PCR mix in cold racks. Each well will now contain 16 µL of material.
   b. Seal the plate with a plastic PCR adhesive seal, centrifuge in a plate spinner for 10 s at 500 × *g*, and pre-amplify libraries by running 8 cycles of the standard PCR program (part 3, point 36).
6. After 8 cycles, centrifuge the plate in a plate spinner for 10 s at 500 × *g* and place in a cold rack.
   a. Transfer the whole volume from 48 wells (half plate) into a fresh 96-well plate for qPCR. Cover the original plate with a plastic PCR adhesive seal and keep it cold.
   b. Add SYBR Green to the wells of the new plate to reach 1× final concentration (Figure 1A).

c. Perform qPCR with the program below. Use an empty well as negative control.

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	1 min	1
Denaturation	98°C	10 s	30 cycles
Annealing	60°C	30 s	
Extension	72°C	30 s	
Hold	4°C	Hold	

d. Observe the qPCR amplification plot. Each curve represents a single well. Select the number of cycles n at which the curves start rising exponentially (Figure 1B).
e. Run n cycles of PCR for the remaining 48 pre-amplified libraries (original plate).

Note: You have now defined the optimal number of PCR cycles to run on single cells from this type of tissue. In future GTAC experiments on the same type of tissue, run 8 + n PCR cycles in total.

7. After PCR, centrifuge the plate in a plate spinner for 10 s at 500 × *g* and pool all the volume from 48 wells into a 15 mL conical centrifuge Falcon tube. The tube will contain ~700 µL of material at this point. Purify the pool:
   a. Add 3500 µL of PB buffer to the pool (5× the volume in the pool). Mix 5–10 times but do not vortex.
   b. Perform a column purification as described in part 4, points 41–44. Elute in 21 µL of EB buffer or nuclease-free water.
   c. Perform a double-sided AMPure XP bead size selection, first with a 0.5× beads:sample ratio, followed by a 1.2× beads:sample ratio, as described in part 4, points 46–55. At the end of bead purification, elute in 10 µL of nuclease-free water.

⚠ CRITICAL: Your sample volume at this stage is half of the volume noted in part 4, point 46. Hence, first add 10.5 µL of beads to the sample, followed by 14.7 µL of beads in the next addition.

8. Quantify the pool using a Qubit dsDNA HS Assay Kit. Using the ng/µL readout, estimate the total mass [ng] of DNA in the pool, and the approximate mass [ng] of DNA per cell.

Note: In our experience, when 16 cycles are performed on human bone marrow $CD34^+$ cells (healthy or leukemic), we obtain ~0.8–3 ng of DNA per single cell. However, the yield will be sample-specific.

9. Check the library size distribution using a microcapillary assay like Agilent Bioanalyzer or TapeStation. Traces indicative of high-quality libraries are shown in Figure 2. Set the region between 100 and 7,000 bp. Calculate the molarity of your library using the average size from the Bioanalyzer or an equivalent assay and the concentration [ng/µL] obtained with Qubit.
   a. Estimate whether the library amount would be sufficient for sequencing, considering that normally you would have 382 cells (~8 times more material) in the pool (the minimum library concentration for sequencing on Illumina platforms is 4 nM).
      i. If the library concentration is sufficient and the fragment size distribution resembles a high-quality ATAC library (Figure 2), use these conditions in the GTAC experiment on the samples of interest.
      ii. If library concentration is insufficient, repeat another test by increasing the number of PCR cycles.

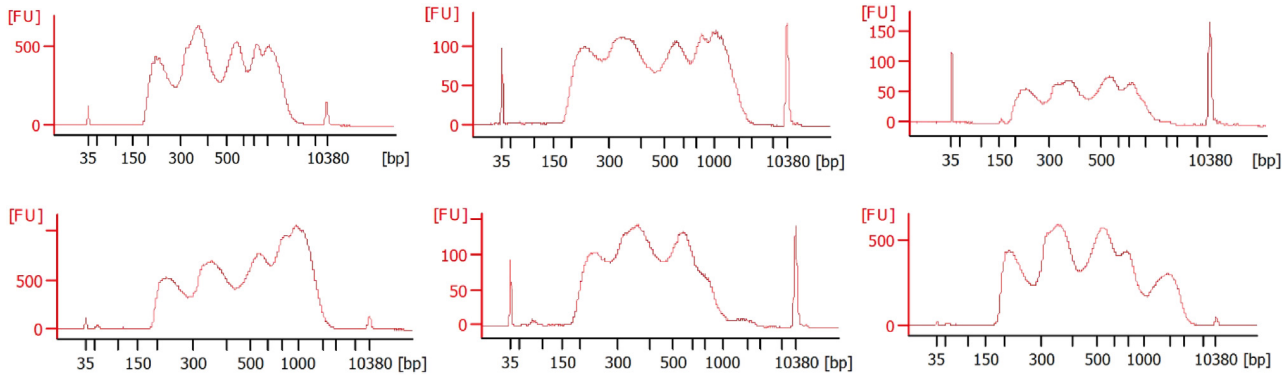


Figure 2. Bioanalyzer traces showing optimal scATAC-seq libraries

For each trace, a variable number of single-cell libraries was pooled and purified. Successful libraries display the characteristic 'wavy' shape, with periodic rises and drops in signal intensity (nucleosomal pattern).

iii. If Bioanalyzer traces are suboptimal, refer to the troubleshooting section for potential causes and solutions.

Preparation two: Design and validation of target-specific pre-amplification and nested genotyping primers using bulk genomic DNA (gDNA)

Timing: 2 days

The main advantage of GTAC over classical scATAC-seq is the amplification of multiple genomic loci harboring mutations, in parallel with amplification of open chromatin fragments.[1] This is achieved by adding target-specific primers (part 3, point 34) to the PCR (part 3, point 36).

Preparation two and preparation three describe how to design and validate these target-specific primers. Two primer pairs are needed for each locus of interest: one used in the first PCR (part 3, point 36), to pre-enrich for the locus – hereafter, we refer to these primers as "pre-amplification genotyping primers"; the second pair is nested within the pre-amplification amplicon and is used in genotyping PCR1 (part 5, points 61–64), to further enrich for the locus, attach plate barcodes, and universal CS1/CS2 adapters – hereafter, we refer to these primers as "nested genotyping primers". A schematic overview is shown in Figure 3.

Note: A list of all pre-amplification and nested genotyping primers that we have previously validated for GTAC applications[1] is provided in Table S1. We advise to use some of these primer pairs as positive controls in points 12 and 13. Additionally, if some of these primers cover the genomic loci of your interest, there is no need to re-design new primers for these loci.

10. Design pre-amplification genotyping primer pairs for each locus using Primer-BLAST[5] or another tool for primer design. We recommend designing at least 2 pairs per locus.
    a. Use the following criteria:
        i. Aim for an amplicon size < 300 bp, if possible (ideally < 200).
        ii. Aim for a primer length between 19 bp and 25 bp.
        iii. Check for primer specificity against a genomic reference. If possible, select those pairs that show no or minimal non-specific binding.
        iv. Melting temperature should be 57°C–63°C; GC content should be 20%–80%.
        v. Once at least 2 primer pairs are designed per locus, use the IDT Oligo Analyzer to check for annealing to generic i7 and i5 primers, for each genotyping primer. Exclude genotyping primers for which the predicted secondary structure of the heterodimer formed with the i7 or i5 primer displays $\Delta G < -9$ kcal/mol, if possible.

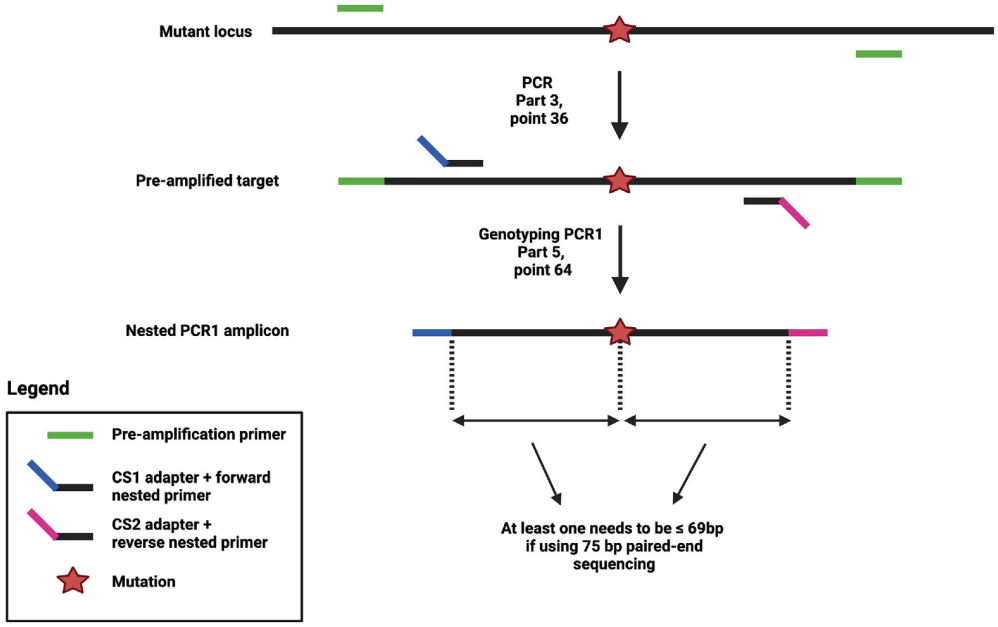


Figure 3. Schematic overview of primers to be tested prior to the GTAC experiment

b. Order the primers and reconstitute them at 100 μM in TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA).
c. Prepare 2 μM primer dilutions to be used in point 12.

Note: Reconstitute the pre-amplification genotyping primers and prepare the 2 μM primer dilutions in a designated pre-PCR clean area, ideally in a biosafety cabinet. The original 100 μM stock, kept in the pre-PCR area, will also be used in preparation three.

Note: Keeping a short amplicon size minimizes the risk of the Tn5 inserting within the amplicon, which may lead to failed genotyping and allelic dropouts, if the targeted locus is found in open chromatin.

⚠ CRITICAL: Pre-amplification genotyping primers for which the predicted secondary structure of the heterodimer formed with the i7 or i5 primer displays $\Delta G < -9$ kcal/mol are likely to interfere with scATAC-seq library generation and should be avoided. Troubleshooting 2.

11. Design nested genotyping primer pairs for each locus using Primer-BLAST[5] or another tool for primer design. We recommend designing at least 2 pairs per locus.
    a. Use the following criteria:
        i. Primers should be nested with respect to those designed in point 10.
        ii. Aim for an amplicon size of 75–200 bp, if possible (not less than 75 bp).
        iii. Design the primers so that the beginning of the forward or of the reverse primer is within 69 bp from the mutation. This is crucial when using 75 bp paired-end sequencing, to efficiently cover the nucleotide of interest. For details about sequencing configurations, refer to the Note after part 5, point 84.
        iv. Aim for primer length between 19 bp and 25 bp.
        v. Check for primer specificity against a genomic reference. If possible, select those pairs that show no or minimal non-specific binding.

vi. Melting temperature should be 57°C–63°C; GC content should be 20%–80%.

b. Order the primers and reconstitute them at 100 μM in TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA). Forward primers should contain the CS1 adapter, while reverse primers should contain the CS2 adapter, at their 5′ ends.

c. Prepare 2 μM primer dilutions to be used in point 13.

12. Validate pre-amplification genotyping primers using bulk gDNA (usually extracted from cell lines):

a. Set up the following reaction for each primer pair:

Pre-amplification genotyping primer testing reaction (bulk gDNA)	
Reagent	1 reaction
Q5 High-Fidelity 2× Master Mix	5 μL
gDNA template 50–150 ng/μL	0.5 μL
Nuclease-free water	3.5 μL
Forward pre-amplification primer 2 μM	0.5 μL
Reverse pre-amplification primer 2 μM	0.5 μL

b. Set up a no template negative control reaction and a positive control reaction with primers that display successful amplification at these conditions (Table S1).

c. Run the following PCR program:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Gap-Filling	72°C	10 min	1
Initial Denaturation	98°C	10 min	1
Denaturation	98°C	10 s	40 cycles
Annealing	60°C	30 s	
Extension	72°C	30 s	
Final extension	72°C	20 s	1
Hold	4°C	Hold	

d. Run the PCR product on a 2% agarose gel and select the pair with the strongest and most specific amplification. Examples are shown in Figure 4.

13. Validate nested genotyping primers using bulk gDNA (usually extracted from cell lines):

a. Set up the following reaction for each primer pair:

Nested genotyping primer testing reaction (bulk gDNA)	
Reagent	1 reaction
KAPA2G Robust HS Ready Mix	3.125 μL
gDNA template 50–150 ng/μL	1.5 μL
Forward genotyping nested primer 2 μM	0.81 μL
Reverse genotyping nested primer 2 μM	0.81 μL

b. Set up a no template negative control reaction and a positive control reaction with primers that display successful amplification at these conditions (Table S1).

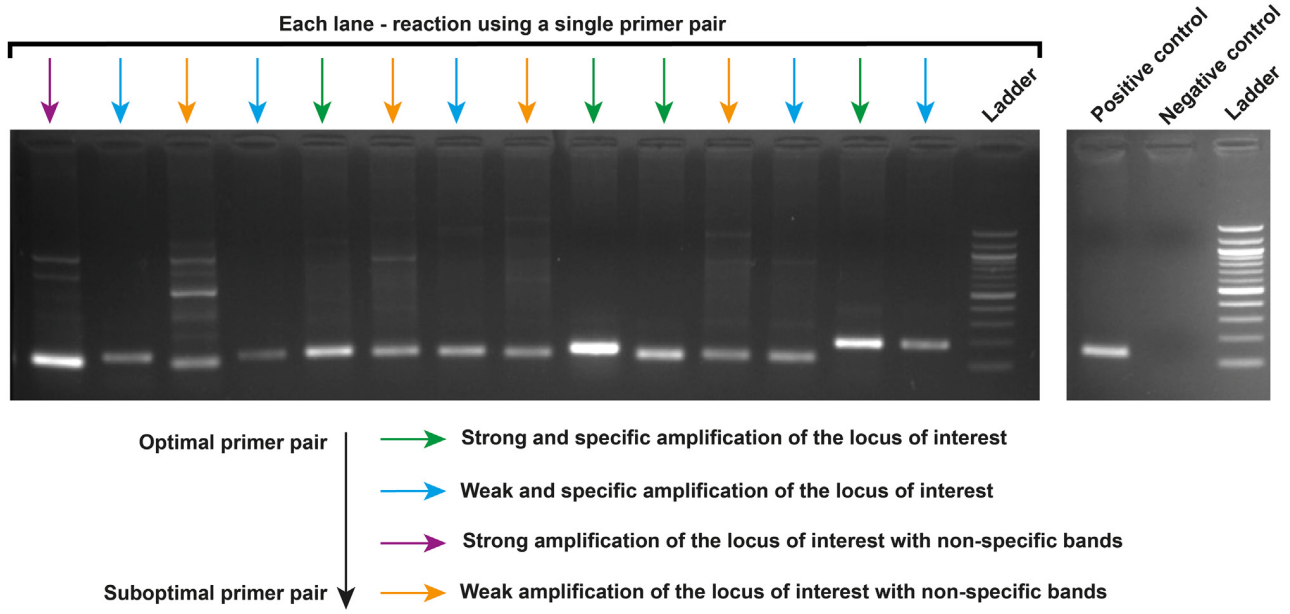


Figure 4. Representative gel of primer testing with bulk gDNA

Left gel: each lane shows the PCR product of a single primer pair used on gDNA. Arrows are color-coded based on the quality of the product. An optimal primer pair displays amplification as indicated by the green arrow. Right gel: a primer pair for which there is prior knowledge of good amplification is used as positive control. A no-template reaction is used as negative control. A 100 bp ladder was used for both gels.

c. Run the following PCR program:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	95°C	15 s	35 cycles
Annealing	60°C	20 s	
Extension	72°C	1 min	
Final extension	72°C	5 min	1
Hold	4°C	Hold	

d. Run the PCR product on a 2% agarose gel and select the pair with the strongest and most specific amplification. Examples are shown in Figure 4. Troubleshooting 11.

Note: For some loci, it may be challenging to obtain strong and specific amplification with bulk gDNA, even when many pairs are tested. In these cases, we still recommend testing primers in single cells, which may show acceptable amplification despite suboptimal results on bulk gDNA.

Note: Given that 2 forward and 2 reverse primers were designed for each locus, we recommend setting up 4 reactions per locus, each testing a different forward/reverse primer combination.

Optional: For some genotyping loci, it may be challenging to design an efficient nested primer pair. In these cases, you may use the same primer pair for the pre-amplification PCR (part 3, point 36) and for PCR1 (part 5, points 61–64). In those cases, test the primer pair

both with the Q5 High-Fidelity 2× Master Mix (point 12) and with the KAPA2G Robust HS Ready Mix (point 13). However, in our experience, genotyping is more efficient when a nested genotyping primer pair is used for PCR1.

Optional: In our applications, we usually genotype point mutations or small indels: hence, the same primer pair amplifies the wild-type (WT) and the mutant alleles. Sometimes, you may need to amplify larger fusion genes – in these cases, you need to design a fusion-specific and a WT-specific primer pair to efficiently capture both alleles. In those cases, remember to test the fusion-specific primer pair on gDNA containing the fusion gene of interest to validate amplification efficiency.

Preparation three: Validation of selected genotyping primers in single cells

Timing: 1–10 days

You now need to test successful pre-amplification and nested genotyping primer pairs selected in preparation two in single cells, to evaluate whether (a) you can efficiently generate genotyping amplicons at single-cell level; (b) scATAC-seq library generation is successful in presence of these primers.

⚠ CRITICAL: If you are genotyping multiple loci in the same sample of interest, pre-amplification genotyping primers for all loci need to be added simultaneously prior to the first PCR (point 15).

Note: For Preparation three, it is not necessary to use samples with specific mutations, as efficiency of mutation detection is not being evaluated at this step. An exception to this is the case in which you are genotyping specific fusion genes (see the consideration at the end of preparation two).

Note: Points 14–15 days (prior to PCR) should be performed in a designated pre-PCR clean area, ideally in a biosafety cabinet.

14. Using one plate sorted in preparation one, point 3, perform lysis and Tween 20 quenching:
    a. Incubate the plate at 65°C for 15 min in a thermal cycler, with the lid heated to 105°C.
    b. Add 2 μL of Tween 20 10% (vol/vol) to each well using an 8 × 20 μL multichannel pipette.
    c. Centrifuge the plate in a plate spinner for 10 s at 500 × *g* and incubate at 21°C for 10 min.
15. Prepare PCR mixes:
    a. Prepare PCR primer mixes with pre-amplification genotyping primers for 30 reactions, according to the table below. The final concentration of each pre-amplification genotyping primer in each well (16 μL) should be 100 nM. Troubleshooting 11.

Note: The number of mixes depends on the number of different conditions you are testing.

Note: The table below shows an example with 2 genotyping loci (4 primers). For more loci, add further primer pairs as needed, in place of water. For > 4 loci (> 8 primers), use a higher initial primer concentration, to keep the volume at 10 μL of mix per reaction.

PCR mix containing pre-amplification genotyping primers

Reagent	1 reaction	30 reactions
Q5 High-Fidelity 2× Master Mix	8 μL	240 μL
Pre-amplification forward primer (locus 1 sample 1) 6.4 μM	0.25 μL	7.5 μL

(*Continued on next page*)

Continued

Reagent	1 reaction	30 reactions
Pre-amplification reverse primer (locus 1 sample 1) 6.4 μM	0.25 μL	7.5 μL
Pre-amplification forward primer (locus 2 sample 1) 6.4 μM	0.25 μL	7.5 μL
Pre-amplification reverse primer (locus 2 sample 1) 6.4 μM	0.25 μL	7.5 μL
Nuclease-free water	1 μL	30 μL

b. Prepare a positive control mix without pre-amplification genotyping primers by mixing 240 μL of Q5 High-Fidelity 2× Master Mix with 60 μL of nuclease-free water.
c. Dispense each mix into 24 wells, by adding 10 μL of mix per well (Figure 5).
d. Seal the plate with a plastic PCR adhesive seal and centrifuge in a plate spinner for 10 s at 500 × *g*.
e. Run the PCR program as in part 3, point 36, by using the number of cycles estimated in preparation one.

16. Aliquot 8 μL of nuclease-free water to a fresh 96-well plate (test genotyping stock plate).
17. When PCR is finished, transfer 4 μL of amplified material to the test genotyping stock plate. Seal the test genotyping stock plate with an aluminum adhesive seal and place it on wet ice.
18. Pool the remaining material from each condition into a separate 2 mL microcentrifuge tube (Figure 5). Each tube should contain ~250 μL of material. Purify the pools as follows:
    a. Add 1250 μL of PB buffer to each pool. Mix 5–10 times but do not vortex.
    b. Perform a column purification as described in part 4, points 41–44. Elute each pool in 21 μL of EB buffer or nuclease-free water.
    c. Perform a double-sided AMPure XP bead size selection, first with a 0.5×, followed by a 1.2× beads:sample ratio, as described in part 4, points 46–55. At the end of bead purification, elute in 8–10 μL of nuclease-free water.

⚠ CRITICAL: Your sample volume is half that in part 4, point 46. Hence, add first 10.5 μL of beads to sample, followed by 14.7 μL of beads in the next addition.

19. Quantify the pool using a Qubit dsDNA HS Assay Kit. Check library size distribution using a microcapillary assay like Agilent Bioanalyzer or TapeStation. Make sure ATAC libraries are comparable to those in Figure 2. Troubleshooting 2.

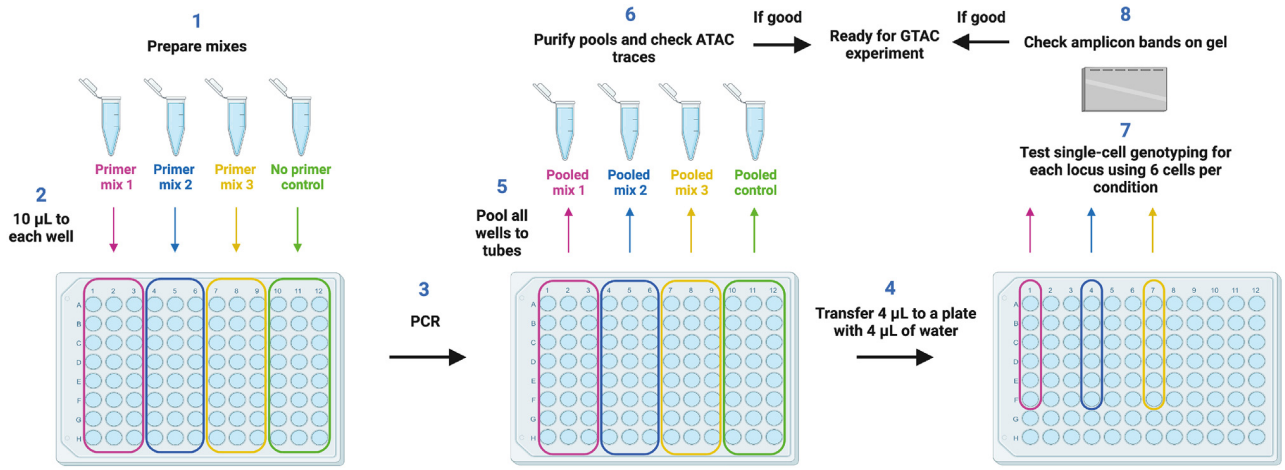


Figure 5. Schematic overview of primer testing in single cells
ATAC traces are tested from pools of 24 single cells. Efficiency of single-cell genotyping is tested individually in single cells. Blue numbers indicate the order in which the steps are executed.

CellPress
OPEN ACCESS

⚠ CRITICAL: If at this point one or more conditions show sub-optimal results, this is likely due to one or more pre-amplification genotyping primer pairs in that condition interfering with ATAC library amplification. In this case, perform another test with only one primer pair per condition, to identify the problematic pair. Re-design primers for this locus and test them as in preparation two and preparation three.

20. Test single-cell genotyping:
    a. Transfer 1.5 μL aliquot from each well of the test genotyping stock plate generated in point 17 as input for the genotyping PCR1 reaction (table below). Use a minimum of 6 wells per condition.

Single-cell genotyping PCR1 testing reaction

Reagent	1 reaction
KAPA2G Robust HS Ready Mix	3.125 μL
Nuclease-free water	1.25 μL
Nested genotyping forward/reverse primer pair 5 μM each primer	0.375 μL
Pre-amplified library	1.5 μL

    b. Run the PCR program from preparation two, point 13c.

Note: Each amplicon needs to be tested in a separate reaction. Hence, if two loci were amplified in one condition in point 15, each well needs to be used as input for two reactions – one amplifying locus 1, and one amplifying locus 2.

21. Run each reaction on a 2% agarose gel to check how many cells display successful amplification, band intensity for each locus, and amplification specificity. Use gDNA as positive control and a no template reaction as negative control. Representative gels in Figure 6.
    a. If primers show acceptable single-cell amplification (ideally at least 5/6 strong and specific bands per locus) and if scATAC-seq library generation is successful, these primers can be used in the GTAC experiment on primary samples. Troubleshooting 11.
    b. Follow preparation two and preparation three to find suitable primers for all loci.

Optional: If more than 4 loci are amplified per cell for a single sample of interest, you need to decide which loci will be amplified in the same genotyping PCR1 reaction (part 5). At the genotyping PCR1 reaction stage, we advise genotyping up to 4 different loci in the same reaction. For example, if you are amplifying 6 loci per cell in a single sample of interest, you should split genotyping PCR1 into two reactions. The choice of which loci to amplify together in the same genotyping PCR1 reaction is based on the results of point 21. We advise combining loci that show strong amplification and keeping them separate from loci showing weaker amplification, to avoid amplification of one locus outcompeting the amplification of the other locus. Troubleshooting 11

Preparation four: Preparation of the primers necessary for GTAC

Timing: 1 h

Once all pre-amplification and nested genotyping primers are validated, you are ready for the execution of GTAC on samples of interest. At this stage, order and prepare all primers necessary for the protocol (sequences listed in the key resources table).

22. Order primers necessary for the preparation of the barcoded lysis buffer (part 1):

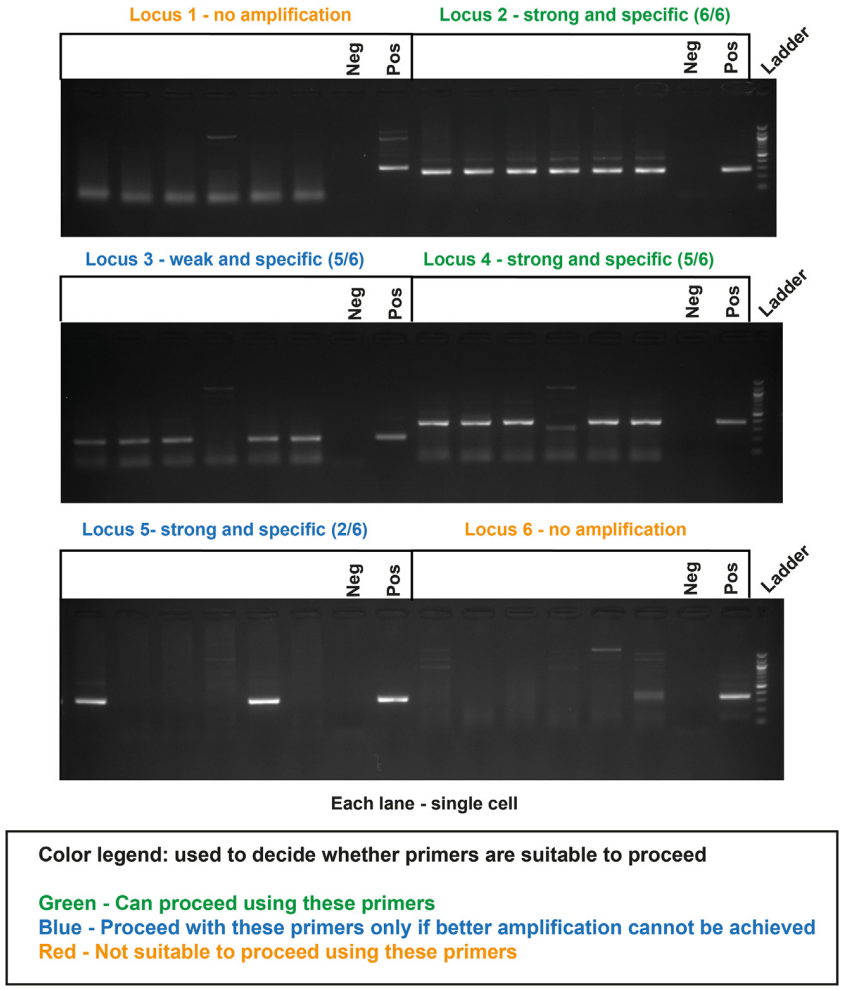


Figure 6. Single-cell genotyping testing results

Primers for each locus were tested in 6 single cells. Each lane represents amplification from a single cell. Depending on the quality and consistency of amplification, primers can be used in a GTAC experiment, or new primers need to be tested for the locus. Bulk gDNA was used as positive control. A no-template reaction was used as negative control.

a. 384 barcoded i7 primers (Table S2) are ordered in plate format, reconstituted at 100 μM in TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA).
   i. Keep the 384 barcoded i7 primers at −80°C prior to use.
b. Reconstitute barcoded i5 primers (Table S2) at 100 μM in TE buffer.
   i. The number of different barcoded i5 primers to order is equivalent to the number of different 384-well plates to be pooled and sequenced on the same sequencing run or on the same lane (for details, refer to part 1).
   ii. Keep the barcoded i5 primers at −20°C prior to use.

23. Order target-specific pre-amplification genotyping primers validated in preparation two and preparation three.
    a. Reconstitute pre-amplification genotyping primers required for part 3 at 100 μM in TE buffer.
    b. Keep reconstituted pre-amplification genotyping primers at −20°C prior to use.
24. Order the Access Array Barcode Library for Illumina Sequencers-384, Single Direction (four 96-well plates containing 40 μL of 2 μM primer pair per well; 384 barcoded primer pairs in total), required for part 5. Hereafter, we refer to these as "barcoded genotyping PCR2 primers".
    a. Prepare two fresh 384-well plates.

b. Transfer 20 μL of 2 μM barcoded genotyping PCR2 primers into each of the two fresh 384-well plates, following these steps:
   i. Aliquot each of the four original 96-well plates into a separate quadrant of the fresh 384-well plate (A1 into quadrant 1, A2 into quadrant 2, etc.).
   ii. Hereafter, we refer to these 384-well plates as "Access Array 2 μM stock plates".
   iii. Keep the Access Array 2 μM stock plates at −20°C prior to use.

Note: Make sure to order sufficient Access Array Barcode Library for Illumina Sequencers-384, Single Direction kits, for the approximate number of plates you plan to sort, considering that each genotyping PCR2 plate (part 5, point 65) requires 1.2 μL of primer pair per well (you might need to prepare multiple Access Array 2 μM stock plates).

25. Order the CS1, LCS1, CS2, and CS2rc custom sequencing primers required for genotyping library sequencing (part 5).
   a. Reconstitute the primers at 100 μM in TE buffer.
   b. Prepare single-use aliquots for each primer:
      i. Primers and volumes required depend on the Illumina sequencing platform you will use to sequence genotyping libraries. Refer to part 5, point 84 for primer volumes required per sequencing run on different Illumina sequencing platforms.
   c. Keep reconstituted single-use aliquots at −20°C prior to use.

⚠ CRITICAL: Handle barcoded i7 and i5 primers, pre-amplification genotyping primers, Access Array 2 μM stock plates, and sequencing primers, in a designated pre-PCR clean area, ideally in a biosafety cabinet, to avoid contamination from PCR products. CS1, CS2, and CS2rc primers contain LNA modifications and should always be aliquoted and stored as single-use aliquots at −80°C.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-human CD45RA (1:40; clone HI100)	BD	Cat# 564552; RRID:AB_2738841
Mouse anti-human CD90 (1:20; clone 5E10)	BioLegend	Cat# 328124; RRID:AB_2561693
Mouse anti-human CD123 (1:40; clone 6H6)	BioLegend	Cat# 306006; RRID:AB_314580
Mouse anti-human CD38 (1:20; clone HIT2)	BioLegend	Cat# 303526; RRID:AB_10983072
Mouse anti-human CD10 (1:40; clone HI10a)	BioLegend	Cat# 312222; RRID:AB_2562157
Mouse anti-human CD34 (1:160; clone 581)	BioLegend	Cat# 343510; RRID:AB_1877153
Mouse anti-human CD117 (1:80; clone 104D2)	BioLegend	Cat# 313238; RRID:AB_2629837
Mouse anti-human CD2 (1:160; clone RPA-2.10)	BioLegend	Cat# 300210; RRID:AB_314034
Mouse anti-human CD3 (1:320; clone HIT3a)	BioLegend	Cat# 300310; RRID:AB_314046
Mouse anti-human CD4 (1:160; clone RPA-T4)	BioLegend	Cat# 300510; RRID:AB_314078
Mouse anti-human CD8a (1:320; clone RPA-T8)	BioLegend	Cat# 301010; RRID:AB_314128
Mouse anti-human CD20 (1:160; clone 2H7)	BioLegend	Cat# 302308; RRID:AB_314256
Mouse anti-human CD235ab (1:320; clone HIR2)	BioLegend	Cat# 306606; RRID:AB_314623
Chemicals, peptides, and recombinant proteins		
Fetal bovine serum, heat inactivated (or equivalent)	Sigma-Aldrich	Cat# F9665-100ML
IMDM, no phenol red	Gibco	Cat# 21056023
DPBS, no calcium, no magnesium	Thermo Fisher Scientific	Cat# 14190169
Bovine serum albumin solution 30% (wt/vol)	Sigma-Aldrich	Cat# A8577
Digitonin	Promega	Cat# G9441
IGEPAL CA-630	Sigma-Aldrich	Cat# I8896
Tween 20 10% (or equivalent polysorbate 20)	Sigma-Aldrich	Cat# 655205-250ML
UltraPure 1 M Tris-HCl, pH 8.0	Thermo Fisher Scientific	Cat# 15568025

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sodium chloride solution 5 M	Sigma-Aldrich	Cat# S6546
$MgCl_2$ (1 M)	Thermo Fisher Scientific	Cat# AM9530G
N,N-dimethylformamide (DMF)	Sigma-Aldrich	Cat# D4551-250ML
Magnesium acetate solution 1 M	Sigma-Aldrich	Cat# 63052-100ML
Potassium acetate solution 5 M	Sigma-Aldrich	Cat# 95843-100ML-F
Illumina Tagment DNA Enzyme and Buffer Small Kit	Illumina	Cat# 20034197
EDTA (0.5 M), pH 8.0, RNase-free	Thermo Fisher Scientific	Cat# AM9260G
SDS solution, molecular biology grade (10% w/v)	Promega	Cat# V6551
DAPI solution (1 mg/mL)	Thermo Fisher Scientific	Cat# 62248
UltraPure agarose	Thermo Fisher Scientific	Cat# 16500500
UltraPure ethidium bromide, 10 mg/mL	Thermo Fisher Scientific	Cat# 15585011
Q5 high-fidelity 2× master mix	New England Biolabs	Cat# M0492L
Nuclease-free water	Thermo Fisher Scientific	Cat# AM9932
Buffer PB, 500 mL binding buffer	QIAGEN	Cat# 19066
Buffer PE (concentrate, 100 mL)	QIAGEN	Cat# 19065
Sodium acetate (3 M), pH 5.5, RNase-free	Thermo Fisher Scientific	Cat# AM9740
KAPA2G Robust HotStart ReadyMix, 6.25 mL	Sigma-Aldrich	Cat# KK5702
FastStart high fidelity PCR system, dNTPack	Sigma-Aldrich	Cat# 4738292001
TE buffer	Thermo Fisher Scientific	Cat# 12090015
Buffer EB (250 mL)	QIAGEN	Cat# 19086
Ethanol absolute	VWR	Cat# 20821.310
Critical commercial assays		
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69506
Qubit dsDNA Quantification, High Sensitivity Assay Kit	Thermo Fisher Scientific	Cat# Q32854
Agilent TapeStation HS D5000 ScreenTape	Agilent	Cat# 5067-5592
Agilent TapeStation HS D5000 Reagents	Agilent	Cat# 5067-5593
Agilent TapeStation HS D1000 ScreenTape	Agilent	Cat# 5067-5583
Agilent TapeStation HS D1000 Reagents	Agilent	Cat# 5067-5584
Agilent High Sensitivity DNA Kit	Agilent	Cat# 5067-4626
Oligonucleotides		
Target-specific genotyping primers (standard desalting)	IDT (designed by the user)	N/A
Target-specific nested barcoded genotyping PCR1 primers (standard desalting)	IDT (designed by the user)	N/A
See Table S2 for 384 indexed i7 primers (standard desalting)	IDT (design: Xu et al.[4])	https://www.nature.com/articles/s41596-021-00583-5
See Table S2 for barcoded i5 primers (HPLC purified)	IDT	N/A
Access Array barcode library for Illumina sequencers-384, single direction, 2 μM	Standard BioTools	Cat# 100-4876
CS1 sequencing primer (HPLC purified): A+CA+CTG+A CGACATGGTTCTACA	IDT	https://www.sciencedirect.com/science/article/pii/S266616672030112X
CS2 sequencing primer (HPLC purified): T+AC+GGT+A GCAGAGACTTGGTCT	IDT	https://www.sciencedirect.com/science/article/pii/S266616672030112X
CS2rc sequencing primer (HPLC purified): A+GAC+CA+ AGTCTCTGCTACCGTA	IDT	https://www.sciencedirect.com/science/article/pii/S266616672030112X
LCS1 sequencing primer (HPLC purified): GGCGACCA CCGAGATCTACACTGACGACATGGTTCTACA	IDT	https://www.sciencedirect.com/science/article/pii/S266616672030112X
Generic i7 primer (standard desalting): CAAGCAGAAG ACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTC GGAGATGT	IDT	N/A
Generic i5 primer (standard desalting): AATGATACGG CGACCACCGAGATCTACACTCGTCGGCAGCGTCA GATGTG	IDT	N/A
Software and algorithms		
Primer-BLAST	Ye et al.[5]	https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-134
TARGET-seq genotyping pipeline	Rodriguez-Meira et al.[6]	https://github.com/albarmeira/TARGET-seq

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bcl2fastq (version 2.20)	Illumina	https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf
Snakemake scATAC-seq pre-processing pipeline	Chen et al.[2]	https://github.com/dbrg77/scATAC_snakemake
R (version 4.1.3)	R-Project	
ArchR (version 1.0.2)	Granja et al.[7]	https://www.nature.com/articles/s41588-021-00790-6
FlowJo (version 9 or 10)	FlowJo	https://www.bdbiosciences.com/en-gb/products/software/flowjo-v10-software
infSCITE	Jahn et al.[8]	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0936-x
Custom scripts for downstream GTAC analysis	Turkalj, Jakobsen et al.[1]	https://github.com/sventurkalj/GTAC/tree/v1.0.0
Other		
12.5 mL GRIPTIP, sterile, filter	INTEGRA Biosciences	Cat# 6455
High volume MANTIS chip	Formulatrix	Cat# MCHVSMR6
Agencourt AMPure XP beads (or equivalent)	Beckman Coulter	Cat# A63881
FrameStar PCR plate 384-well, skirted	Azenta Life Sciences	Cat# 4ti-0384/C
FrameStar 96-well semi-skirted PCR plate	Azenta Life Sciences	Cat# 4ti-0900/C
MicroAmp optical 96-well reaction plate	Thermo Fisher Scientific	Cat# N8010560
Axygen 96-well clear V-bottom 500 µL polypropylene deep well plate	Corning	Cat# P-96-450V-C
Adhesive PCR plate seals (plastic)	Thermo Fisher Scientific	Cat# AB0558
Self-adhesive plate seal, aluminum, thick 60 µm	Starlab (UK) Ltd	Cat# E2796-0792
qPCR adhesive seal	Azenta Life Sciences	Cat# 4ti-0560
20 mL extender tube (or equivalent)	Angen Biotech	Cat# D50071
QIAvac 24 Plus (or equivalent)	QIAGEN	Cat# 19413
Falcon 5 mL round bottom polystyrene test tube, with cell strainer snap cap (or equivalent)	Corning	Cat# 352235
QIAquick spin column	QIAGEN	Cat# 28104
DNA LoBind tube 1.5 mL	Eppendorf	Cat# 022431021
NucleoCounter NC-3000 (or equivalent)	ChemoMetec	Cat# 991-3001
NC-slide A8 (or equivalent)	ChemoMetec	Cat# 942-0003
Solution 13 (or equivalent)	ChemoMetec	Cat# 910-3013
25 mL disposable reagent reservoir, sterile (or equivalent)	Corning	Cat# RES-V-25-S
CoolRack XT PCR384 thermoconductive tube rack for 384-well PCR plates (or equivalent)	Azenta Life Sciences	Cat# BCS-538
ThermoMixer C (or equivalent)	Thermo Fisher Scientific	Cat# 15158953
Centrifuge 5430 R (or equivalent)	Eppendorf	Cat# 5428000655
Centrifuge 5910 R (or equivalent)	Eppendorf	Cat# 5943000061
MPS 1000 mini plate spinner (or equivalent)	Labnet	Cat# C1000
ProFlex 96-well PCR system (or equivalent)	Thermo Fisher Scientific	Cat# 4484075
ProFlex 384-well PCR system (or equivalent)	Thermo Fisher Scientific	Cat# 4484077
QuantStudio 3 real-time PCR system, 96-well, 0.2 mL, laptop (or equivalent)	Thermo Fisher Scientific	Cat# A28567
Invitrogen Qubit 3 fluorometer (or equivalent)	Thermo Fisher Scientific	Cat# 15387293
Qubit assay tubes (or equivalent)	Thermo Fisher Scientific	Cat# Q32856
2100 bioanalyzer instrument	Agilent	Cat# G2939BA
4200 TapeStation system (or equivalent)	Agilent	Cat# G2991BA
Magnetic stand-96 (or equivalent)	Thermo Fisher Scientific	Cat# AM10027
Magnetic separation rack, 0.2 mL tubes (or equivalent)	EpiCypher	Cat# 10-0008
GelDoc Go gel imaging system (or equivalent)	Bio-Rad	Cat# 12009077
PowerPac basic power supply (or equivalent)	Bio-Rad	Cat# 1645050
Fisherbrand Midi Plus horizontal gel system (or equivalent)	Thermo Fisher Scientific	Cat# 11833293
MA900 multi-application cell sorter (or equivalent)	Sony	N/A
MANTIS automated liquid dispenser (or equivalent)	Formulatrix	N/A
Mosquito HTS nanolitre liquid handler (or equivalent)	SPT Labtech	N/A
VIAFLO 96/384 electronic pipette (or equivalent)	Integra Biosciences	N/A

MATERIALS AND EQUIPMENT

Note: FACS buffer and thawing media recipes listed below apply to thawing and staining procedures involving human cryopreserved bone marrow/peripheral blood mononuclear cells.

FACS buffer	
Reagent	Amount
Fetal Bovine Serum (FBS)	50 mL
DNase I (10 mg/mL)	500 μL
IMDM (no phenol red)	450 mL
Total	**500.5 mL**
Store at 4°C for a maximum of two weeks.	

Thawing media for cryopreserved human bone marrow/peripheral blood mononuclear cells	
Reagent	Amount
Fetal Bovine Serum (FBS)	5 mL
DNase I (10 mg/mL)	500 μL
FACS buffer	45 mL
Total	**50.5 mL**
Store at 4°C for a maximum of two weeks.	

Note: Pass the FACS buffer and thawing media through a sterile filter prior to storage.

Note: Prepare FACS buffer without DNase I (below) only if FACS staining and sample pre-sorting prior to Omni-ATAC are required. Label it appropriately to distinguish from normal FACS buffer.

- FACS buffer without DNase I: mix 50 mL of FBS with 450 mL of IMDM (no phenol red). Cells will be pre-sorted into this buffer prior to Omni-ATAC.

Store at 4°C for a maximum of 1 month.

- Prepare DPBS with 0.5% BSA by mixing 500 μL of BSA 10% (wt/vol) with 9.5 mL of DPBS 1×.

Store at -20°C for a maximum of 1 month.

Note: Regardless of the starting tissue type, prepare the buffers according to the recipes below, originally provided by Xu et al., 2021[4] and Xu et al., 2022.[9]

Nucleus dilution buffer		
Reagent	Final concentration	Amount
BSA 10% (wt/vol)	0.5%	500 μL
DPBS 1×	0.475×	4.75 mL
Nuclease-free water	N/A	4.75 mL
Total	N/A	**10 mL**
Store at -20°C for a maximum of 1 month.		

Omni-RSB buffer		
Reagent	Final concentration	Amount
Tris-HCl 1 M pH 8	10 mM	500 µL
NaCl 5 M	10 mM	100 µL
$MgCl_2$ 1 M	3 mM	150 µL
Nuclease-free water	N/A	49.25 mL
Total	**N/A**	**50 mL**
Store at 4°C for a maximum of two months.		

4× THS-TD buffer		
Reagent	Final concentration	Amount
Tris-HCl 1 M pH 8	132 mM	132 µL
Potassium acetate 5 M	264 mM	52.8 µL
Magnesium acetate 1 M	40 mM	40 µL
N,N-Dimethylformamide (DMF)	64%	640 µL
Nuclease-free water	N/A	135.2 µL
Total	**N/A**	**1 mL**
Store at −20°C for a maximum of one month.		

⚠ CRITICAL: Handle DMF under a fume hood.

2× tagmentation stop buffer		
Reagent	Final concentration	Amount
Tris-HCl 1 M pH 8	10 mM	100 µL
EDTA 0.5 M	20 mM	400 µL
BSA 10% (wt/vol)	2%	2 mL
Nuclease-free water	N/A	7.5 mL
Total	**N/A**	**10 mL**
Store at −20°C for a maximum of two months.		

STEP-BY-STEP METHOD DETAILS

Part 1: Barcoded lysis buffer preparation

Timing: 3–5 h

In this section, we describe how to prepare 384-well plates containing the fully barcoded lysis buffer, which is done one day before the single nucleus sort (Figure 7). At the end of the section, each well will contain one of the 384 uniquely barcoded i7 primers (well indexes) and each plate will contain a different plate-specific barcoded i5 primer (plate index). i7 and i5 primers amplify and fully barcode tagmented fragments of open chromatin (scATAC-seq fragments).

⚠ CRITICAL: Decide in advance how many plates will be sequenced on the same sequencing run (or in the same lane), to estimate the required number of different i5 indexes. Refer to the note after part 4, point 60, for considerations about the required sequencing depth per cell for scATAC-seq libraries.

Optional: Here, we describe the procedure for generating five lysis buffer 384-well plates, but more plates can be prepared at the same time for convenience and stored at −80°C for at least 2 months.

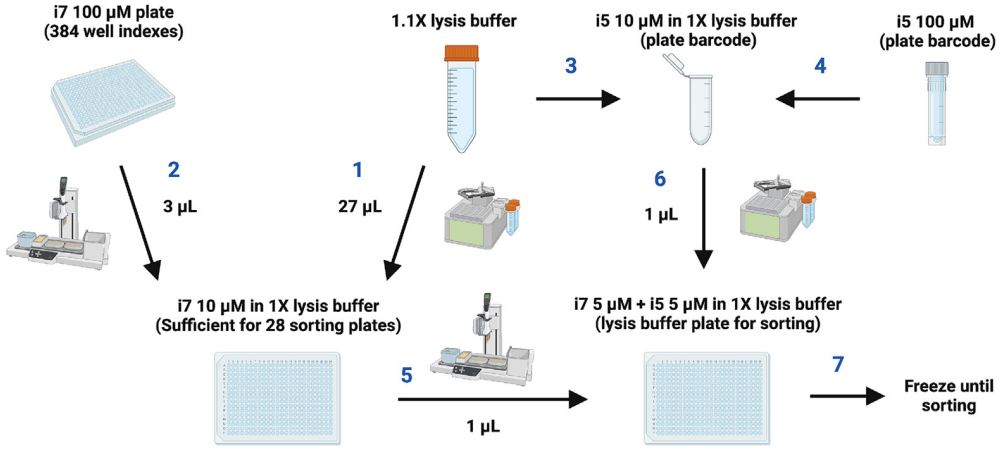


Figure 7. The automated preparation of barcoded lysis buffer

Liquid handling platforms used at individual steps, as well as volumes transferred, are indicated next to the arrows. Blue numbers indicate the order in which the steps are executed.

Note: 384-well or 96-well plates can be used. Here, we describe the procedure for 384-well plates. If 96-well plates are used instead, we recommend doubling all the volumes.

Note: Part 1 should be performed in a designated pre-PCR clean area, ideally in a biosafety cabinet, to avoid contaminations from PCR products.

Note: In the following steps, we describe the usage of the MANTIS and of the INTEGRA VIAFLO 96/384 electronic pipettor in detail. These liquid handling platforms will also be used throughout the later stages, but for detailed instructions about their setup and usage please refer to these initial steps. We advise getting familiar with the working procedures of these platforms in advance (Figures 8 and 9 and Methods videos S1, S2, S3, and S4). Note that the use of these platforms is not essential, but drastically increases throughput. Finally, note that alternative equivalent liquid handling platforms can be used.

1. Prepare 50 mL of 1.1× lysis buffer as outlined in the table below:

1.1× lysis buffer	
Reagent	Amount
Tris-HCl 1 M pH 8	550 µL
NaCl 5 M	110 µL
SDS 10% (wt/vol)	1100 µL
Nuclease-free water	48.25 mL

   a. Mix well using a serological pipette. Keep the buffer cold.

2. Dispense 27 µL of 1.1× lysis buffer into each well of a fresh 384-well plate.

Note: Here, we describe the procedure using the MANTIS liquid handling platform (alternative platforms can be used).

   a. Switch on the MANTIS.
      i. Clean the MANTIS plate holder and chip holder with 70% ethanol.

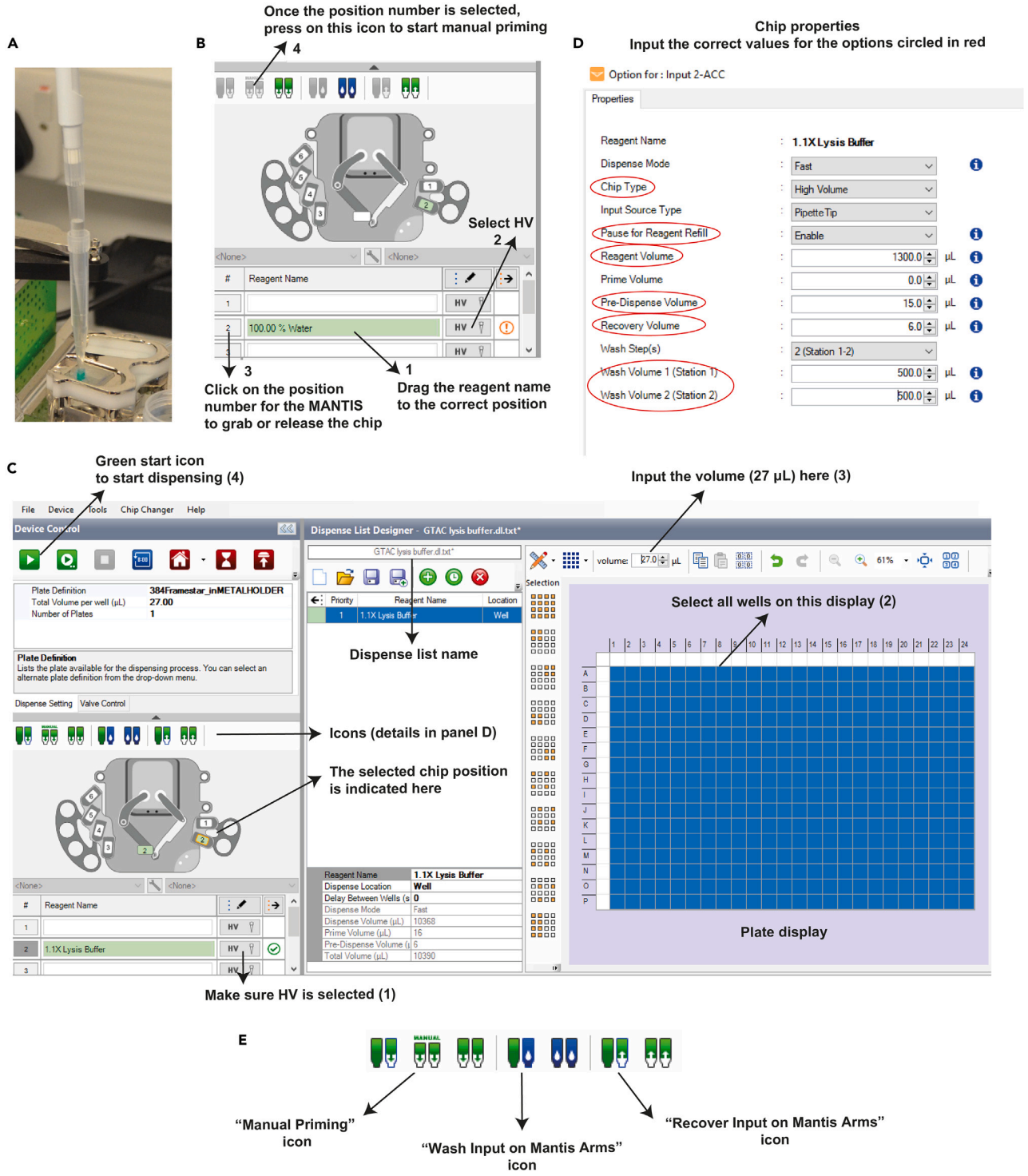


Figure 8. MANTIS operating instructions

(A) Pipetting liquid into a non-filtered tip inserted into the HV chip.

(B) Steps for setting up manual priming with water. Numbers indicate the order in which steps are executed.

(C) Dispense list display. Numbers in brackets indicate the order in which steps are executed.

(D) Chip properties display. Red circles indicate the parameters that need to be defined by the user.

(E) Icons used for priming, washing, and volume recovery from the chip.

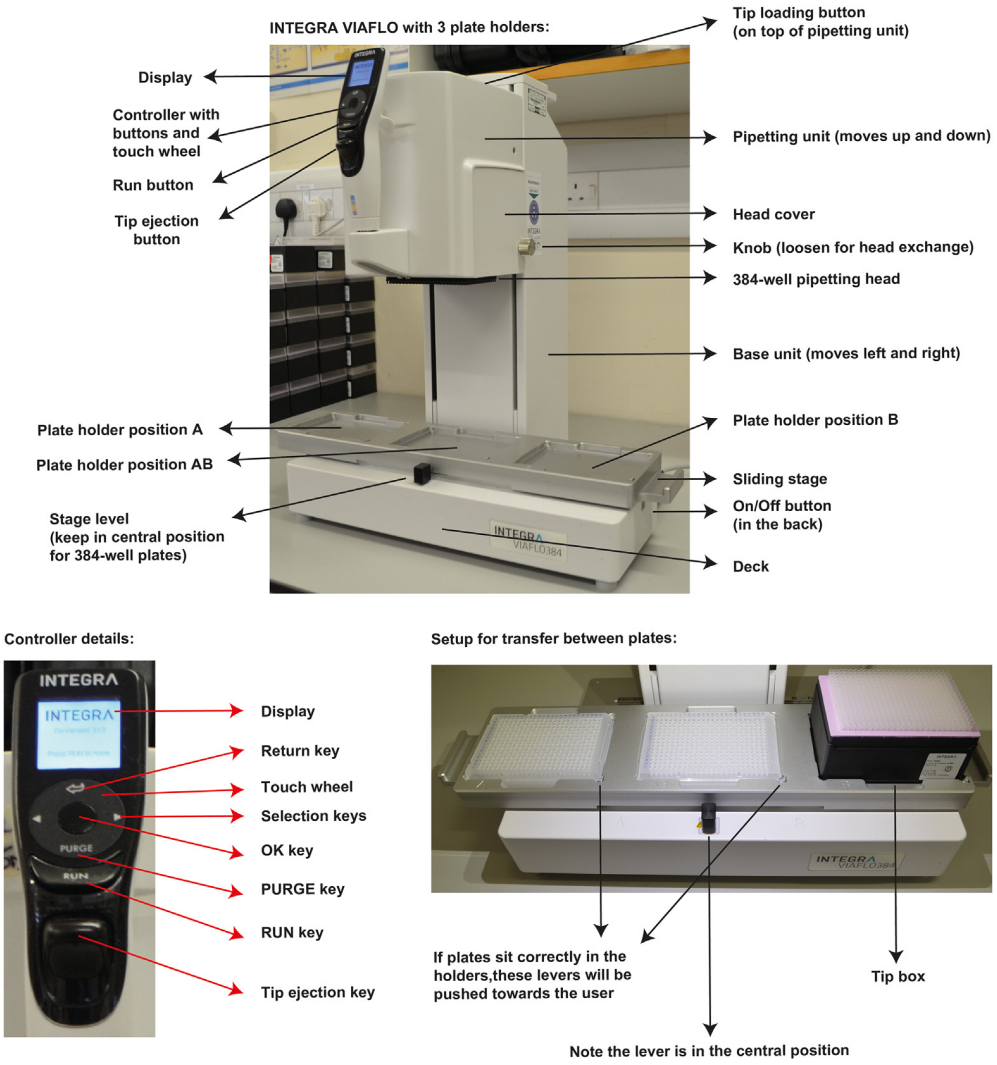


Figure 9. INTEGRA VIAFLO components

Top: components of the VIAFLO. Bottom left: components of the VIAFLO controller. Bottom right: setup for volume transfer between plates. Note that this instrument has 3 plate holders, but an instrument with 2 plate holders can also be used.

ii. Open the MANTIS software.
iii. Insert a High Volume (HV) chip into the chip holder (Methods video S1).

b. Perform manual priming of the chip:
i. Insert a 1 mL non-filtered tip into the chip.
ii. Pipette 500 µL of nuclease-free water into the tip (Figure 8A).
iii. Make sure "HV" is selected next to the "Reagent Name" column (Figure 8B).
iv. Click on the chip position number: this indicates to the MANTIS arm to grab the chip at the indicated position (Figure 8B).
v. Click on the "Manual Priming Input on MANTIS Arms" icon (Figure 8B) and wait until all water passes through the chip. When there is no liquid coming out, stop the manual priming.
vi. Click the position number again to release the chip, which is now ready to be used for lysis buffer dispensing.

vii. Remove the tip from the chip (Methods video S1).

c. Create a new Dispense List (Figure 8C) and name it appropriately. Make sure the program is calibrated for the appropriate plate type and height of the plate rack.
   i. Insert "GTAC 1.1× lysis buffer" as name of your reagent.
   ii. In the plate display, select all wells and input 27 µL in the volume window above.
   iii. Drag the name of your reagent to the "Reagent Name" column (Figure 8C).
   iv. Click HV to the left to specify the chip options. In the menu, set "Pre-dispense" to 15 µL; "Recovery" to 6 µL; "Enable pause for reagent refill"; "Tip volume" to 1300 µL; "Wash Station 1" and "Wash Station 2" both to 500 µL (Figure 8D).

d. Place a 384-well cold rack onto the MANTIS plate holder and place a fresh 384-well plate into the rack. Make sure the plate sits well inside the rack.
   i. Insert a new non-filtered 1 mL tip into the chip.
   ii. Pipette ~1500 µL into the tip. Ensure that there are no bubbles at the bottom of the tip.
   iii. Click the green start icon (Figure 8C) to start dispensing (Methods video S1). Depending on the "Tip volume" you set, the dispense will pause after a given time and allow you to refill the tip with 1.1× lysis buffer.

e. Once the dispense is done, remove the plate rack from the MANTIS, cover the plate with an aluminum or plastic adhesive seal, and centrifuge in a plate spinner for 10 s at 500 × *g*. You can leave the plate at 21°C.

f. Replace the 1 mL tip on the chip with an empty clean tip for washing.
   i. Click the chip position number to grab the chip.
   ii. Click the "Wash Input on MANTIS Arms" icon (Figure 8E) – this will flush the chip with 500 µL of 70% ethanol and 500 µL of water.
   iii. When done, click the "Recover Input on MANTIS Arms" icon (Figure 8E) 1–2 times to aspirate any remaining volume from the chip.
   iv. Click on the chip position again to release the chip.
   v. Remove the tip.

Note: We perform the water priming step (point 2b) each time we start using the MANTIS. Hereafter, we refer to this step as 'manual priming with water'.

Note: We recommend using one MANTIS HV chip exclusively for lysis buffer preparation steps.

⚠ CRITICAL: Set a high enough "Pre-dispense" volume (at least 15 µL) to get air bubbles outside of the chip prior to dispensing. This avoids insufficient volume being dispensed inside the first wells. Hence, whenever using the MANTIS, prepare at least ~15% of dead volume of the dispensed mix.

⚠ CRITICAL: Make sure to have enough 70% ethanol and nuclease-free water in the washing tubes to properly flush the chip. Always use nuclease-free water.

3. Transfer 3 µL of the 100 µM barcoded i7 indexes to the 1.1× lysis buffer plate, to obtain a stock plate of 10 µM barcoded i7 indexes in 1× lysis buffer.

Note: We perform this step using the INTEGRA VIAFLO electronic pipettor with a 384-well 12.5 µL pipetting head (Figure 9), but this step can be performed with any multichannel pipette system.

a. Switch on the VIAFLO. When prompted, press the RUN key to home the instrument, and when prompted press the RUN key again to home the pipettor.

b. Make sure the 384-well pipetting head is installed. If you need to switch heads:
   i. Scroll to "Toolbox" on the display with the touch wheel, select it with the OK key, and select "Change Head" with the OK key.
   ii. Follow the instructions on the display (Methods video S2).
c. Select "Pipet" from the display. Set the pipetting volume to 3 μL, Aspirate Speed to 4, and Dispense Speed to 1.
d. Load the tips (Methods video S3):
   i. Make sure the stage lever is set to the central position (Figure 9).
   ii. Place a new 384 tip box in the plate holder position B.
   iii. By holding the controller, lower the pipetting unit towards the tips until the tip loading button starts flashing.
   iv. Push the tip loading button while keeping the pipetting unit down. When the tip loading button stops flashing, pull up the pipetting unit. Tips should now be fully loaded on the head.
e. Place the 100 μM barcoded i7 index plate and the 1.1× lysis buffer plate in two adjacent plate holders (positions A and AB, respectively).
f. Insert the tips close to the bottom of the 100 μM barcoded i7 index plate and aspirate 3 μL by pressing the RUN key on the controller.
   i. Move the pipetting unit up and insert the tips into the 1.1× lysis buffer plate.
   ii. Press the RUN key again to release the volume and wait until the display signals "Done". This will generate a 10 μM barcoded i7 index plate at a total volume of 30 μL per well (Methods video S4).
   iii. Change the pipetting volume to 12 μL on the display while the tips are still inserted in the plate.
   iv. Aspirate and dispense using the RUN key 3–5 times to properly mix the i7 indexes with the lysis buffer (Methods video S4).
g. Once mixing is complete:
   i. Remove the 100 μM barcoded i7 index plate from the holder, cover well with two aluminum adhesive seals, and place on dry ice for snap freezing.
   ii. Remove the stock 10 μM i7 index plate, cover with an aluminum or plastic adhesive seal, and centrifuge in a plate spinner for 20 s at 500 × *g*.
   iii. Label the plate as "i7 10 μM stock" and let it rest for 10 min at 21°C.
   iv. Place the plate in the fridge or on a cold rack.
h. Finally, remove the tips from the pipetting head: place the empty 384 tip box on a plate holder position B, insert the tips into the box (do not touch the bottom of the box), and press the tip ejection key twice (Methods video S3). Dispose of the tips.

⚠ CRITICAL: Take absolute care to avoid any cross-well contamination of i7 indexes at this point, especially when handling the 100 μM barcoded i7 index plate.

Note: When dispensing the 384 barcodes, visually inspect the volume in the tips to make sure all the barcodes were aspirated.

Note: When using the VIAFLO to transfer or mix viscous solutions (as the 1.1× lysis buffer), keep the Dispense Speed low, to avoid retention of the volume inside the tips.

4. Dilute the 100 μM i5 plate barcodes in 1.1× lysis buffer:
   a. Mix 50 μL of 100 μM i5 plate barcode with 450 μL of 1.1× lysis buffer, to obtain a 10 μM i5 plate barcode/1× lysis buffer mix.
   b. Mix well and keep cold.
   c. Repeat for all i5 barcodes.

⚠ CRITICAL: Avoid any cross-contamination of different i5 barcodes. Clean the pipettes between dispenses with 70% ethanol, to avoid traces of the previous i5 barcode

contaminating the next one. Also, make sure to use filtered tips! If you notice cross-contamination, do not use these barcodes!

Note: The number of different 10 μM i5 mixes depends on the number of 384-well plates that will be multiplexed on the same sequencing run. We usually multiplex 5 plates on the same sequencing run; hence, we prepare 5 different 10 μM i5 plate barcode mixes.

Note: The volumes here are sufficient to prepare one 384-well plate per i5 barcode. If you are preparing more 384-well plates with the same i5 barcode, prepare more 10 μM i5 plate barcode/1× lysis buffer mix accordingly.

5. Transfer 1 μL of the 10 μM barcoded i7 indexes from the i7 10 μM stock plate into fresh 384-well plates. Prepare five fresh 384-well plates.

Note: Here, we describe the procedure using the VIAFLO in Repeat Dispense pipetting mode.

a. Load a new set of 384 tips (Methods video S3).
b. Place the i7 10 μM stock plate and the first fresh 384-well plate on two plate holders of the VIAFLO. Keep the stage lever at the central position.
c. In the display, select "Repeat Dispense". Set "Dispense" to 1 μL; "Count" to 6; "Aspiration Speed" to 1; "Dispense Speed" to 1.
d. Aspirate 6 μL from the i7 10 μM stock plate.
e. Transfer 1 μL into the fresh 384-well plate (insert the tips close to the well bottom).
f. Remove the aliquoted plate from the holder and cover with a plastic or aluminum adhesive seal.
g. Repeat points 5e and 5f for the remaining fresh 384-well plates.
h. Dispense the remaining volume into the i7 10 μM stock plate with the PURGE key.
i. Remove the tips (Methods video S3).
j. Centrifuge all the plates in a plate spinner for 10 s at 500 × *g*. Snap freeze the i7 10 μM stock plate on dry ice. Place the five aliquoted 384-well plates in the fridge. Switch off the VIAFLO.

⚠ CRITICAL: When aliquoting small volumes into empty plates as described here, make sure that the tips are in close proximity of the well bottom. Otherwise, the liquid may be retained on the tips.

⚠ CRITICAL: After all plates are dispensed, visually inspect the bottom of each plate to make sure that each well contains the i7 index. Sometimes, a few wells are not dispensed due to liquid retention on the tip. If this happens, manually transfer 1 μL from the i7 10 μM stock plate to the missing wells.

6. Finally, aliquot 1 μL of 10 μM i5 index (point 4) into the five plates containing 1 μL of 10 μM i7 indexes (point 5), to obtain 2 μL of fully barcoded lysis buffer per well. A different 10 μM i5 index will be aliquoted into each plate.

Note: Here, we describe the procedure using the MANTIS.

a. Create a new Dispense List to aliquot 1 μL of into each well. Re-use the HV chip from point 2.
b. Place a new non-filtered 1 mL tip into the chip and transfer 500 μL of the first 10 μM i5 index into the tip.
c. Place the first 10 μM i7 index plate on the cold rack onto the MANTIS plate holder.
d. Aliquot 1 μL of 10 μM i5 index into each well. This will generate a 5 μM i7/i5 indexed 1× lysis buffer plate used for sorting (final volume of 2 μL per well).
e. When the dispense is done, remove the cold rack from the plate holder.

i. Label the plate with the name of the appropriate i5 index.
ii. Seal carefully with an aluminum adhesive seal.
iii. Centrifuge in a plate spinner for 20 s at 500 × *g* and freeze on dry ice.

f. Replace the tip on the HV chip and wash the HV chip as in point 2f.
g. Carefully clean the HV chip on the outside:
   i. Remove the HV chip from the holder.
   ii. Wipe the chip on the bottom side with paper inserted in 70% ethanol.
   iii. Clean the HV chip with dry paper. This removes traces of index contamination.
h. Repeat points 6a–6g for all plates, using different 10 μM i5 indexes.
   i. Wash the HV chip, store it in a 50 mL conical centrifuge Falcon tube, and switch off the MANTIS.

⚠ CRITICAL: Always wash and clean the HV chip properly between different i5 indexes, to avoid contaminating plates with the wrong plate index.

⚠ CRITICAL: When freezing plates, always use aluminum instead of plastic adhesive seals.

Optional: If two people are working, one person can perform point 5 and the other person can simultaneously perform point 6. This setup will speed up the execution of these steps.

Pause point: You can store the frozen 5 μM i5/i7 indexed 1× lysis buffer plates at −80°C until the sort or, if you prepare more plates than needed, for at least 6 months.

Part 2: Sample preparation, pre-sorting, tagmentation, and single-nucleus sorting

Timing: 1 day

In this section, we describe the protocol for sample processing, pre-sorting of populations of interest, Omni-ATAC on cell populations, and single-nucleus sorting of tagmented nuclei into 5 μM i5/i7 indexed 1× lysis buffer plates prepared in part 1.

Note: In our applications, usually only a fraction of cells from the initial sample is of interest for GTAC analysis. Hence, we first pre-sort cells of interest into a tube, perform Omni-ATAC on these cells, and only then sort single tagmented nuclei. Points 8–17 describe in detail our custom thawing and FACS staining procedures for pre-sorting, optimized for cryopreserved human bone marrow (BM)/peripheral blood (PB) hematopoietic cells, adapted from Rodriguez-Meira et al., 2020.[6] An example of 2× antibody mixes and 2× FMO control mixes that we use to stain human BM mononuclear cells can be found in Table S3.

7. If you are working with cryopreserved human BM or PB mononuclear cells and need to pre-sort cell population of interest prior to Omni-ATAC using FACS, proceed directly to point 8. For other scenarios and/or tissues, follow one of the points below, depending on sample type:
   a. If you work with a different tissue type (fresh or cryopreserved), just a fraction of the sample is of interest, and you use FACS to pre-enrich for the desired cell fraction.
      i. Follow in-house protocols for tissue dissociation (and/or thawing) and FACS staining (instead of following points 8–18).
      ii. After staining, proceed with point 19 (pre-sorting).
   b. If just a fraction of the sample (fresh or cryopreserved) is of interest and you use a pre-enrichment strategy different from FACS sorting:
      i. Follow in-house protocols for tissue dissociation (and/or thawing) and pre-enrichment (instead of following points 8–22).
      ii. Proceed directly with point 23 (preparation for Omni-ATAC).

c. If the whole sample (fresh or cryopreserved) is of interest (hence, any kind of pre-enrichment is not necessary prior to Omni-ATAC):
   i. Follow in-house protocols for tissue dissociation (and/or thawing) followed by nuclei extraction (instead of following points 8–24f).
   ii. Proceed directly with point 24 g (nuclei centrifugation and tagmentation).

Note: Prior to point 24g, for optimal GTAC results, you need to have 20,000–100,000 nuclei per microcentrifuge tube. < 20,000 cells might result in fewer sorted plates, while > 100,000 cells may lead to suboptimal tagmentation. Supplementary Protocol 2 (points 1–14) from Corces et al., 2017[10] provides guidelines on nuclei isolation from solid cryopreserved tissues, prior to Omni-ATAC.

d. If you are working with cultured cells and pre-enrichment is not necessary:
   i. Transfer 20,000–100,000 cells into a microcentrifuge tube and start at point 23 (instead of following points 8–22). Cell viability should ideally be > 85% prior to Omni-ATAC procedures.

Note: If cell viability is lower, refer to Supplementary Protocol 1 from Corces et al., 2017,[10] for guidelines on how to avoid suboptimal results in case of low viability.

Optional: Regardless of the sample preparation procedures, if cell numbers are high, you can split cells into multiple microcentrifuge tubes with 20,000–100,000 cells per tube prior to Omni-ATAC and run several tagmentation reactions in parallel.

⚠ CRITICAL: To have appropriate single-cell genotyping controls and to control for batch effects in scATAC-seq, we sort cells from a WT sample into one column of each plate. Hence, in parallel with the sample of interest, we always process a cryopreserved control sample too. This control sample is also used for FMOs and sorting gate settings. Make sure not to cross-contaminate samples.

⚠ CRITICAL: During all steps of sample preparation, pipette mix gently not to lyse fragile cells. Perform points 9–18 under a sterile laminar flow cabinet and use filtered tips.

8. Prepare for cryopreserved sample processing:
   a. Prepare a box of wet ice and place FACS buffer on ice.
   b. Bring cryopreserved samples from liquid nitrogen on dry ice.
   c. Set the water bath to 37°C and warm the thawing media and FBS.
   d. Set up the FACS sorter (we use the Sony MA900 Multi-Application Cell Sorter, but other machines can be used).
9. Thaw the samples. Work with a maximum of two samples at a time:
   a. Place samples in the water bath and wait until they are 70%–80% liquid. Proceed when there is still a small piece of frozen tissue inside the cryovial.
   b. Dropwise, add 1 mL of warm FBS to the sample using a P1000.
   c. Pipette mix very gently and transfer all the volume into a 15 mL conical centrifuge Falcon tube.
   d. Dropwise, add 1 mL of warm thawing media to the tube. Hand-mix gently.
   e. Wash the cryovial with 1 mL of thawing media and add dropwise to the tube.
   f. Slowly, add 6 mL of thawing media to the tube to reach a total of 10 mL. After every 2 mL added, hand-mix gently.
   g. Place the tube on ice.
10. When all samples are ready, centrifuge at 350 × *g* for 10 min. Remove supernatant, ideally by pouring it gently and then blotting the tube on dry paper, to remove traces of liquid.
11. Resuspend cells in 1 mL of ice-cold FACS buffer and mix gently.

a. Pass cells through a 35 μm filtered FACS tube.
b. Wash the 15 mL tube with 1 mL of FACS buffer. If starting cell number was > 20 million, add an additional 3 mL of FACS buffer, for a total of 5 mL.

12. Count cells either with trypan blue and a hemocytometer or using an automatic cell counter.
    a. Take note of total cell numbers and viability.
    b. Based on the number of FMOs used, set aside an aliquot of control cells, and place them on ice.
13. Centrifuge other samples at 350 × *g* for 5 min. Remove supernatant as in point 10. ~50 μL of supernatant will remain in the tube.
14. Work without light. Add 2× antibody full-stain mixes to each sample. We add 50 μL of antibody mix if cell number is < 10 million or 100 μL of antibody mix for higher cell numbers.

Place samples at 4°C, in the dark, and incubate for 30–40 min.

Note: Some cells clump during this step and cannot be properly resuspended in the antibody mix. We tend to eliminate these clumps with the pipette tip, given that they may lead to sub-optimal staining.

15. Take the control cell aliquot set aside in point 12b and mix gently.
    a. Add cells to each 2× FMO mix and mix gently.
    b. Place FMO tubes at 4°C, in the dark.
    c. Incubate for 30–40 min.
16. Prepare single-staining controls used to set the correct voltages for each fluorochrome.

Note: To save precious cells, we use commercial beads for single-staining controls. Ideally, control cells should be used for single stains, to more closely match sample fluorescence characteristics.

17. When incubation is complete, wash samples and FMOs with 1 mL of ice-cold FACS buffer and centrifuge at 350 × *g* for 5 min at 8°C.
    a. During centrifugation, mix 10 mL of ice-cold FACS buffer with 1 μL of DAPI to obtain FACS buffer with $1/10^4$ DAPI.
18. Remove supernatant as in point 10.
    a. Resuspend samples in 700 μL-1 mL of FACS buffer with $1/10^4$ DAPI.
    b. Resuspend FMOs in 100 μL of FACS buffer with $1/10^4$ DAPI, apart from the FMO for the DAPI channel, which should be resuspended in 100 μL of FACS buffer.
    c. Keep cells cold, in the dark.

Optional: You can use a nuclear dye other than DAPI. In that case, it is crucial to keep the same dye for successive single-nucleus sorting! Refer to Xu et al., 2021[4] for a benchmark of several dyes.

Note: For samples with high cell numbers or displaying visible clumping, repeat cell filtering.

19. Set up the pre-sorting FACS panel:
    a. Use single stains to set up voltages.
    b. Run each FMO control and perform "Manual Compensation" to optimize the compensation matrix.
    c. Using a combination of FMO readouts and the full-stained control sample, set up the sorting gates.
20. Check tube sorting calibration (Figure 10A):
    a. Simulate a sort using sheath fluid, to collect a few droplets of sheath into test tubes.

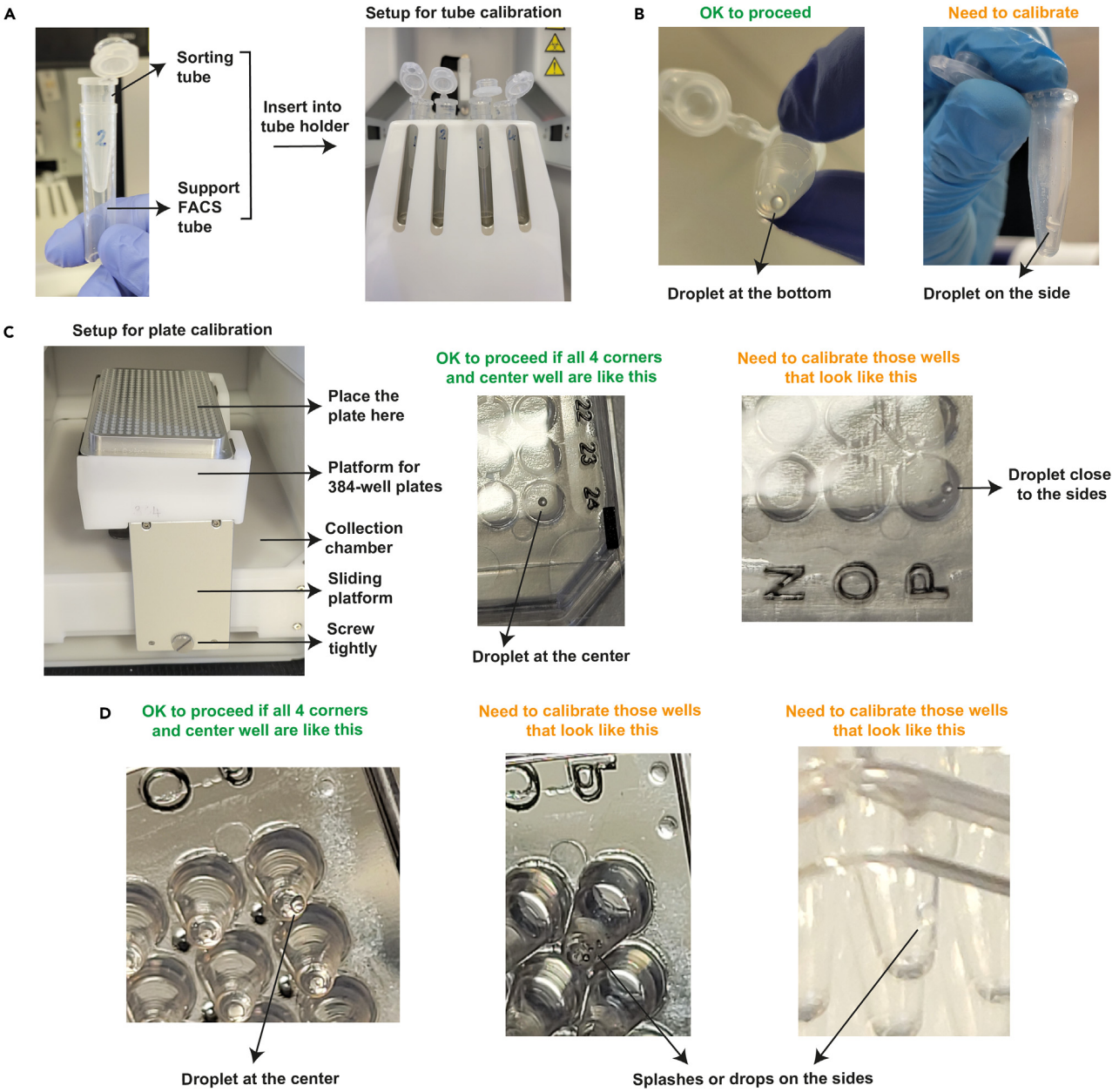


Figure 10. Sony MA900 cell sorter calibration

(A–D) Panels A and B provide the visual guidelines for calibrating the sorter for sorting into microcentrifuge tubes. Panels C and D provide the visual guidelines for calibrating the sorter for sorting into 384-well plates.

b. Calibrate the streams so that the droplets of sheath fluid are collected at the bottom of the tubes and not on the sides (Figure 10B).

21. Put 200 µL of FACS buffer without DNase I in 1.5 mL microcentrifuge tubes to which cells will be sorted. Place tubes on ice. Set the collection and sample chambers to 4°C.

⚠ CRITICAL: To minimize cell loss, pre-coat tubes with 1 mL of FACS buffer without DNase I for 5 min.

⚠ CRITICAL: At this point, populations of interest will be sorted into tubes. Each tube will subsequently be subject to Omni-ATAC, which erases most surface protein information due to membrane lysis. If you wish to retain FACS population information for successive single-nucleus sorting, sort different populations into different tubes and label tubes according to the population sorted.

⚠ CRITICAL: Aim to sort 20,000–100,000 cells into each tube. Keep different samples in separate tubes.

22. Sort single, live cells of interest into microcentrifuge tubes with 200 μL of FACS buffer without DNase I:
    a. Set the sorting mode according to cell numbers and desired purity.
    b. After the sort is finished, place tubes on ice, in the dark.
23. After sorting, proceed immediately with the following steps:
    a. Pre-chill a fixed-angle centrifuge to 4°C and heat the thermomixer to 37°C.
    b. Check the viability of the sorted samples. If the viability is < 85%, we advise to consider the indications by Corces et al., 2017.[10]
    c. Prepare digitonin 0.1% by mixing 1 μL of digitonin 2% (wt/vol) with 19 μL of nuclease-free water and keep on ice.
    d. Prepare Omni-RSB-T buffer by mixing 10 mL of Omni-RSB buffer with 100 μL of Tween 20 10% (vol/vol) and keep on ice.
    e. Prepare the Omni-RSB-DTN buffer and the tagmentation mix (without the Tn5 enzyme) according to the tables below and keep on ice.

Omni-RSB-DTN buffer

Reagent	Volume
Omni-RSB buffer	487.5 μL
Tween 20 10% (vol/vol)	5 μL
IGEPAL CA-630 10% (vol/vol)	5 μL
Digitonin 2% (wt/vol)	2.5 μL

Tagmentation mix (no Tn5 enzyme)

Reagent	Volume for 1 sample
4× THS TD buffer	12.5 μL
Digitonin 0.1% (wt/vol)	5 μL
Nuclease-free water	27.5 μL

Note: Avoid freezing and thawing digitonin 2% more than three times. Prepare aliquots of digitonin 2% in advance, prior to storage at −20°C.

⚠ CRITICAL: Work quickly and keep your samples on ice. Always keep track of the microcentrifuge tube orientation in the fixed-angle centrifuge. For very small cells or for low cell numbers, it may not always be possible to see the cell pellet. It is crucial not to disturb pellets when aspirating supernatants, which drastically reduces the nuclei yield prior to sorting. Work next to a light source, to carefully observe any movement of cell pellets.

24. Perform Omni-ATAC on the sorted samples following the steps below:
    a. Centrifuge samples at 500 × *g* for 5 min at 4°C in the pre-chilled fixed-angle centrifuge.
    b. Carefully aspirate the supernatant. Use two pipetting steps: collect most supernatant with a P1000 tip and then collect the remaining liquid with a P200 pipette, by fitting a non-filtered 20 μL tip on top of a 200 μL tip. Keep the bottom of the tip on the opposite side of the pellet.

c. Resuspend the pellet in 1 mL of cold DPBS with 0.5% BSA and centrifuge samples at 500 × *g* for 5 min at 4°C in the pre-chilled fixed-angle centrifuge.
d. Remove 900 μL of supernatant with a P1000 tip and the remaining 100 μL as in point 24b.
e. Resuspend cells in 50 μL of Omni-RSB-DTN buffer. Pipette gently up to 6 times and leave on ice for exactly 3 min. Troubleshooting 5 and 8.
f. Add 1 mL of Omni-RSB-T buffer to dilute the lysing detergents. Invert the tube 3 times gently with your hand.
g. Centrifuge nuclei at 1,000 × *g* for 5 min at 4°C in the pre-chilled fixed-angle centrifuge. During centrifugation, for each sample, add 5 μL of Tn5 enzyme to 45 μL of tagmentation mix, pipette mix 10 times, and keep on ice.
h. Remove 950 μL of supernatant with a P1000 tip and the remaining 100 μL as in point 24b. Take care not to disturb the pellet. Troubleshooting 7 and 10.

Optional: If cell numbers are low, Omni-ATAC may result in severe cell loss and sorting multiple plates may not be feasible. In these cases, an alternative strategy is to perform FAST-ATAC instead of Omni-ATAC. For details about FAST-ATAC procedures prior to single nuclei sorting, refer to the detailed guide by Xu et al., 2021[4] (specifically, point 6B). However, note that this may result in lower scATAC-seq library complexity and higher percentage of mitochondrial reads.

⚠ CRITICAL: Considerable nuclei loss can occur at this step. The nuclei pellet is usually harder to see than a cellular pellet and, sometimes, a spread of nuclei can be observed instead of a round pellet. If pellets start falling toward the bottom of the tube, leave a few μL on the bottom. However, do not leave excessive liquid in the tube – this causes Tn5 dilution and retention of mitochondrial DNA.

Note: When working with many samples, pellets drop slowly toward the bottom of the tube. This is usually not a point of concern if you can still see them, but make sure not to aspirate pellets.

25. Resuspend nuclei in 50 μL of Tn5-tagmentation mix and pipette mix gently ~20 times. Place tubes on the pre-heated thermomixer and incubate at 37°C for 30 min with 800 rpm shaking.
    a. During incubation, thaw the 5 μM i5/i7 indexed 1× lysis buffer plates (part 1) at 21°C.
    b. Prepare a box of dry ice next to the sorter.
26. After 30 min, add 50 μL of 2× tagmentation stop buffer to the nuclei and mix gently up to 10 times. Keep on ice for ≥ 10 min. Nuclei can stay in this buffer for ~3–4 h. Troubleshooting 4.
27. Proceed immediately with sorter calibration for 384-well plate sorting:
    a. Cover an empty 384-well plate (same model as the ones you will be sorting nuclei into) with a plastic PCR adhesive seal and insert into the collection platform.
    b. Select 384-well plate sorting:
        i. In "Sort Settings", select "Plate Adjustment".
        ii. Sort 30 droplets of sheath fluid on positions A1, A24, P1, P24, and I13 ("Four Corners and Centre Well").
        iii. Adjust the droplet position and repeat the process iteratively to center all the droplets in the middle of the wells (Figure 10C).
    c. Unseal the plate and repeat point 27b by collecting the droplets inside the wells. If needed, adjust the positions iteratively until the droplets are collected to the bottom of each well (Figure 10D). Make sure there is no splashing on the sides of the wells.
    d. Finally, select the "One Fourth of All Wells" option to sort 30 droplets of sheath fluid into every fourth well. Check the positions of most droplets are optimal (Figure 10D). Troubleshooting 1, 3, and 9.

⚠ CRITICAL: Improper calibration will result in empty wells.

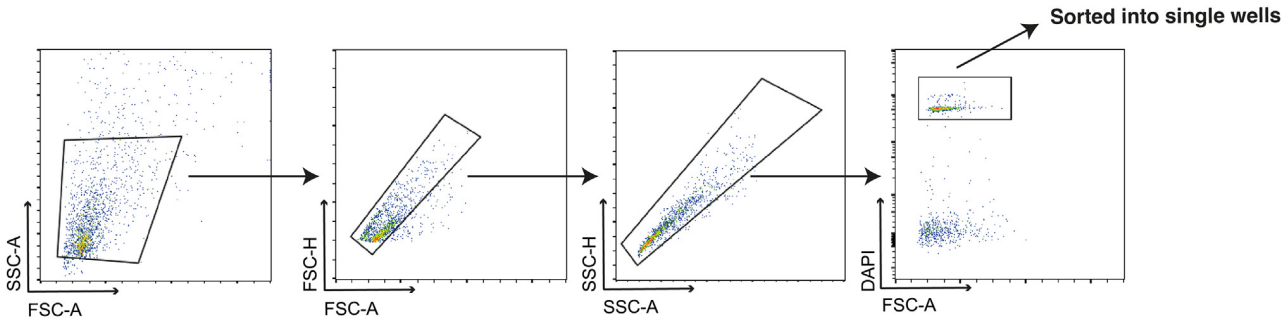


Figure 11. FACS gating strategy for single-nucleus sorting

Optional: For safety, calibration can be re-checked using the control sample instead of the sheath fluid. We recommend doing that when already satisfied with sheath droplet positions.

28. Add 500–900 μL of nucleus dilution buffer to all samples, mix gently 10–20 times, and transfer all the volume into FACS tubes.
    a. Run each sample for a few seconds at lowest pressure to observe the event rate for each sample.
    b. Dilute samples enough to obtain an event rate of 5–10 events/s. Add more nucleus dilution buffer to samples accordingly.
29. Run the control sample for a few seconds and record events, to define gates for nuclei, singlets, and DAPI-negative events (Figure 11).
    a. Add 1 μL of DAPI to 99 μL of nucleus dilution buffer (DAPI 1/100).
    b. Add DAPI 1/100 to the control sample for a final DAPI concentration of $1/10^4$. Mix gently and wait for 5–10 min.
    c. Run the control sample to define the DAPI-positive gate (Figure 11). Troubleshooting 9.
    d. Stain all samples for a final $1/10^4$ DAPI concentration.

Note: We recommend stringency when setting the DAPI-positive gate, to capture only fully lysed nuclei. Two consecutive singlet gates should also be stringent enough to remove most clumps.

⚠ CRITICAL: After adding DAPI, mix gently and wait for ~10 min for full nuclear staining to take place.

30. Centrifuge all the 5 μM i5/i7 indexed 1× lysis buffer plates in a plate spinner for 20 s at 500 × *g*. Carefully unseal the first plate and proceed to single-nucleus sorting:
    a. Keep the event rate < 10 events/s, if possible.
    b. Sort DAPI-positive events into all the columns. Sort WT control cells into ~10% of the wells (we usually use column 24). Keep two wells as empty controls.

Note: We use a sorting mode in which 50% of the droplet before and after the target droplet must be empty, and in which the target event needs to be within the 75% central part of the droplet.

Optional: Using a more stringent sorting mode will result in fewer doublets but fewer sorted plates.

⚠ CRITICAL: Avoid splashes when unsealing lysis buffer plates to prevent cross-well contamination!

31. Once finished, cover the plate carefully with an aluminum adhesive seal, centrifuge in a plate spinner for 20 s at 500 × *g*, and snap freeze on dry ice. Repeat points 30–31 for all plates.

Pause point: Sorted plates can be stored at −80°C for up to 3 months. Tagmented DNA is still not gap-filled at this stage, however. It is hence ideal to start with the next steps as soon as possible.

Part 3: Plate processing, PCR, and genotyping stock preparation

Timing: 2–3 h

In this section, we describe how to process the sorted 384-well plates. After Tn5 release and Tween 20 quenching, each plate undergoes a PCR reaction in order to amplify tagmented fragments of open chromatin. Crucially, target-specific pre-amplification genotyping primers validated in preparation two and preparation three are added to the PCR reaction to pre-enrich for genomic loci of interest. After PCR, a portion of the material will be transferred to a new genotyping stock plate which will be used for single-cell genotyping library construction (part 5). The remaining material will be pooled in order to purify the fully barcoded scATAC-seq libraries (part 4).

Note: Points 32–35 should be performed in a designated pre-PCR clean area, ideally in a biosafety cabinet, prior to PCR.

32. Thaw plates at 21°C. Incubate plates at 65°C for 15 min in a thermal cycler, with the lid heated to 105°C. This step releases the Tn5 from tagmented DNA.
33. Aliquot 1 µL of Tween 20 10% (vol/vol) into each well, to quench SDS:

Note: We perform this step using the MANTIS liquid handler platform (for details on MANTIS usage refer to part 1) and using an HV chip designated for GTAC pre-PCR steps.

a. Manually prime the MANTIS with 500 µL of nuclease-free water.
b. Unseal the plate (avoiding splashes) and aliquot 1 µL of Tween 20 10% (vol/vol) into each well.
c. When a plate is aliquoted, seal it with an aluminum or plastic adhesive seal, centrifuge in a plate spinner for 10 s at 500 × *g*, and leave at 21°C for 10 min.
d. Repeat points 33a–33c for all plates.
e. After all plates are aliquoted, wash the HV chip.

34. Prepare the PCR mix as detailed in the table below. Mix well and keep on a cold rack.

PCR mix containing pre-amplification genotyping primers

Reagent	1 reaction	800 reactions (2 plates)
Q5 High-Fidelity 2× Master Mix	4 µL	3200 µL
Nuclease-free water	0.5 µL	400 µL
Forward primer 1 6.4 µM	0.125 µL	100 µL
Reverse primer 1 6.4 µM	0.125 µL	100 µL
Forward primer 2 6.4 µM	0.125 µL	100 µL
Reverse primer 2 6.4 µM	0.125 µL	100 µL

Note: This mix contains target-specific primers (100 nM final concentration per primer) required for initial amplification of genotyping loci, simultaneously with amplification of tagmented DNA. Note that here, we show an example with 2 genotyping loci (4 primers). For more loci, add further primer pairs as needed, replacing the equivalent volume of water.

For > 4 loci (> 8 primers), use a higher initial primer concentration, to keep the volume of mix at 5 μL per reaction.

⚠ CRITICAL: Make sure all genotyping primers were appropriately tested prior to this step.

Optional: Alternatively, you can use the NEBNext High-Fidelity 2× PCR Master Mix (optimized for NGS libraries; cat #M0541L), but make sure that primers were tested with the enzyme you are using at this stage.

Optional: We use a final concentration of 100 nM per genotyping primer. However, in case of suboptimal genotyping, this concentration may be increased up to 400 nM, with prior testing (preparation three).

35. Using the MANTIS, dispense 5 μL of PCR mix into each well of each 384-well plate.
    a. When a plate is aliquoted, seal well with a plastic PCR adhesive seal, centrifuge in a plate spinner for 10 s at 500 × *g*, and place on wet ice.
    b. When all plates are aliquoted, wash the HV chip, place it in a 50 mL conical centrifuge Falcon tube, and switch off the MANTIS. Troubleshooting 2.

Note: If aliquoting the same PCR mix into multiple plates, there is no need to wash the HV chip between plates. However, if different mixes are aliquoted, wash the HV chip between plates.

⚠ CRITICAL: Use the plate sealer to tightly seal the plate with a plastic PCR adhesive seal, to avoid evaporation. For all the following points, work outside the pre-PCR clean area.

36. Outside of the pre-PCR clean area, incubate the plates using the following PCR program: troubleshooting 1, 3, 4, and 11.

PCR cycling conditions

Steps	Temperature	Time	Cycles
Gap-Filling	72°C	10 min	1
Initial Denaturation	98°C	1 min	1
Denaturation	98°C	10 s	Cycles defined in preparation one (we use 16–18)
Annealing	60°C	30 s	
Extension	72°C	30 s	
Final extension	72°C	5 min	1
Hold	4°C	Hold	

⚠ CRITICAL: The PCR will fail if the first 72°C gap-filling step is not performed.

37. While the PCR is running, prepare the number of fresh 384-well plates corresponding to the number of plates undergoing PCR.
    a. Aliquot 4 μL of nuclease-free water into every well of each plate using the MANTIS.
    b. Label each plate "genotyping stock". Seal the plates.
38. Transfer 2 μL of amplified material from the PCR plate into the genotyping stock plate.

Note: Here, we describe the procedure using the VIAFLO.

    a. When PCR is finished, centrifuge all the plates in a plate spinner for 20 s at 500 × *g*.
    b. Carefully unseal the first plate and place in a plate holder of the VIAFLO.

c. Place one genotyping stock plate in another plate holder, with the stage lever set to the central position.
d. Transfer 2 μL and mix a few times.
e. Seal the genotyping stock plate, label appropriately, centrifuge in a plate spinner for 10 s at 500 × *g*, and snap freeze on dry ice.
f. Repeat points 38a–38e for all plates, changing tips each time.
g. Store genotyping stock plates at −80°C.

⚠ CRITICAL: scATAC-seq libraries are fully barcoded at this stage but genotyping amplicons are not yet barcoded. Avoid any cross-well contamination while preparing genotyping stock plates.

Pause point: Genotyping stocks, used in part 5, can be stored at −80°C for 6 months.

Part 4: Single-cell ATAC-seq library pooling, purification, and preparation for sequencing

Timing: 3–4 h

In this section, we describe how to pool and purify scATAC-seq libraries from the same plate, perform quality control (QC) using capillary electrophoresis, and how to pool libraries from different plates at equimolar ratios to prepare the final scATAC-seq library for sequencing. At this point, scATAC-seq libraries are fully barcoded, due to the unique combination of i7 well index and i5 plate index. Prior to QC, each plate is pooled into a different tube.

⚠ CRITICAL: Incubate an aliquot of AMPure XP beads at 21°C for at least 30 min. These beads will be used in point 46.

39. Pool one plate at a time following these steps:
    a. Centrifuge plates containing amplified scATAC-seq libraries in a plate spinner for 10 s at 500 × *g*.
    b. Pool all the material (~6 μL) from each well using a multichannel pipette:
        i. Pool the material first into a strip of 8 PCR tubes.
        ii. When all the wells of a single plate are pooled in the strip, transfer all the volume from the strip into a 15 mL conical centrifuge Falcon tube.
    c. Repeat points 39a–39b for each plate, pooling each plate into a separate tube. You should have ~ 2 mL of material in each tube.

Optional: For a quicker pooling strategy, refer to step 23 of the protocol by Xu et al., 2021.[4]

40. Add 10 mL of PB buffer (5 × sample volume) to each tube. Mix 5–10 times with the serological pipette. Do not vortex.

Optional: Add 10–30 μL of sodium acetate pH 5.5 to each pool to increase DNA adsorption.

41. Install fresh QIAquick columns on a vacuum and install a 20 mL extender tube on top of each column (Figure 12). Label extender tubes and columns with the appropriate plate name. Load each pool through a column and wait until the entire volume has passed through the column.
42. Wash each column with 5–10 mL of PE wash buffer and wait until the entire volume has passed through the column.

⚠ CRITICAL: Remember to add ethanol to a new bottle of PE buffer and label the bottle in advance.

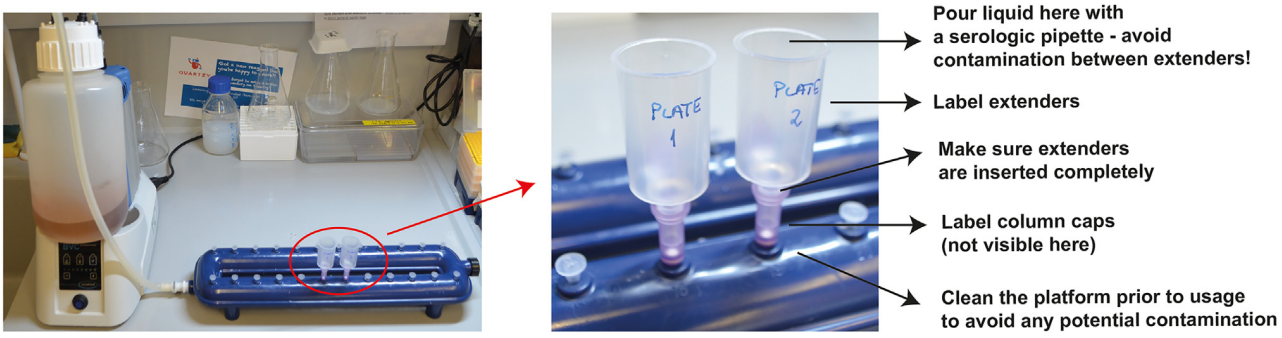


Figure 12. Setup for scATAC-seq library purification via QIAquick columns with extender tubes

43. Remove extender tubes from the columns. Remove columns from the vacuum and put each column into a collection tube. Centrifuge at 18,000 × *g* for 1 min to remove traces of PE buffer.
44. Discard collection tubes and place columns in fresh DNA LoBind 1.5 mL microcentrifuge tubes (or equivalent), labeled on the sides.
    a. Pipette 21 µL of EB buffer directly on the membrane of each column, without touching the membrane.
    b. Incubate at 21°C for 1 min.
    c. Centrifuge at 18,000 × *g* for 1 min.
45. Repeat the elution with 21 µL of EB buffer (point 44) to maximize DNA yield. At the end of this step, each microcentrifuge tube will contain 42 µL of material.
46. Vortex AMPure XP beads thoroughly.
    a. Add 21 µL of beads (0.5× beads:sample ratio) to each pool and mix 20–30 times. Avoid bubbles.
    b. Incubate at 21°C for at least 5 min – this will allow large unwanted DNA fragments to bind the beads.
    c. During incubation, prepare fresh 80% (vol/vol) ethanol by mixing 24 mL of ethanol 100% with 6 mL of nuclease-free water.

Note: Depending on magnet type, bead purification can be performed in microcentrifuge tubes, PCR tubes, or on a V-bottom 96-well plate. The choice will usually depend on the number of pools to purify. For > 8 pools, we use the V-bottom 96-well plate and a plate magnet.

47. Place tubes on a magnetic stand. Beads will accumulate towards the magnet. Wait for the liquid to become completely clear of beads (minimum 3 min). At this point, unwanted large fragments of DNA are bound to the beads, while fragments of interest are in the supernatant.
48. When the supernatant is completely devoid of beads:
    a. Transfer the whole 63 µL of supernatant into new tubes, labeled appropriately.
    b. Add 29.4 µL of AMPure XP beads to each pool and mix 20–30 times. Avoid bubbles.
    c. Incubate at 21°C for at least 5 min.
49. Place the tubes on a magnetic stand and wait for the liquid to become completely clear of beads (minimum 3 min). This time, the fragments of interest are bound to the beads. Points 50–52 must be performed with the tubes placed on the magnetic stand all the time.
50. Carefully remove and discard all supernatant (~93 µL), taking care not to aspirate any beads.
51. Proceed immediately with 80% ethanol washing:
    a. Dropwise, add 200 µL of 80% ethanol to each pool, not disturbing the beads.
    b. Incubate for 30 s.
    c. Remove and discard all 80% ethanol, not disturbing the beads.
    d. Repeat points 51a–51c 1–2 more times.
    e. After the final wash, carefully remove any trace of ethanol using 20 µL tips.

52. Let the beads air-dry for a few minutes (usually 3–5 min, depending on the volume of the beads). The beads are dry enough when the surface of the pellet changes from shiny to matt. While waiting, pre-aliquot nuclease-free water or EB buffer to PCR strips.

⚠ CRITICAL: If the bead pellet starts cracking, proceed to point 53 immediately. If beads over-dry, DNA yield will be reduced, and beads will be challenging to resuspend. However, it is important to dry them sufficiently, given that traces of ethanol could interfere with subsequent QC steps.

53. Remove tubes from the magnet. Resuspend beads in 22 µL of nuclease-free water or EB buffer. Mix thoroughly 20–30 times to resuspend the beads and incubate at 21°C for 3 min.
54. Place the tube back on the magnet and wait for 2–3 min for the supernatant to be completely clear of beads. The DNA fragments of interest are now in the supernatant.
55. Transfer 22 µL of clear supernatant to fresh DNA LoBind 1.5 mL microcentrifuge tubes (or equivalent), labeled appropriately. Take care not to transfer any beads. Troubleshooting 6.
56. Measure the concentration of each pool using the Qubit dsDNA HS Assay Kit. Use 1 µL of each library. Take note of the pool concentration [ng/µL].
57. For each pool, check scATAC-seq library size distribution using a microcapillary assay such as Agilent Bioanalyzer. If pool concentration [ng/µL] is > 20 ng/µL, dilute it prior to running the Bioanalyzer. Check average library size by choosing the region between 100 and 7,000 bp. Representative good quality libraries can be found in Figure 2. A "wavy" shape is expected, with variable peak heights.
    a. If you observe excessive traces of primers or primer dimers, perform an additional AMPure XP bead purification using 1.2× bead:sample ratio.
    b. If a large high-bp peak (details in Figure 18) is present, repeat the AMPure XP double-side exclusion (points 46–55).

Note: If running many pools, we recommend using either the Agilent 4200 TapeStation System with a High Sensitivity D5000 Kit in 96-well plate format, or the Agilent Fragment Analyzer HS NGS Fragment Analysis Kit (1–6000 bp).

58. Once you obtain the concentration [ng/µL] and the average size for each pool, combine equimolar ratios of each pool that will be sequenced on the same run into a fresh DNA LoBind 1.5 mL microcentrifuge tube (or equivalent). Use Table S4 as pooling template.
59. Quantify the final pool using the Qubit dsDNA HS Assay Kit and estimate its average size with Agilent Bioanalyzer or TapeStation.
    a. Input the results into appropriate columns in Table S4 to estimate library molarity.
    b. Dilute the library to 4 nM for sequencing on Illumina platforms.

⚠ CRITICAL: Be precise when diluting the final library for sequencing. Excessive library will result in flow cell over-clustering, which might cause the sequencing run to fail. If the library is too diluted, under-clustering will result in low data amount, and excessive under-clustering may lead to a failed run.

60. Sequence the libraries on an Illumina platform.

Note: The choice of platform and kit depends on the desired sequencing depth. We usually aim for 200,000 reads/cell, and thus sequence 5 pooled 384-well plates on an Illumina NextSeq platform, using a High Output kit. If more plates are sequenced together, a HiSeq or a NovaSeq platform might be more suitable. We usually perform 75 bp paired-end sequencing (150 cycles), but it is also possible to sequence using fewer cycles. Classic Illumina primers from the kits are used for sequencing.

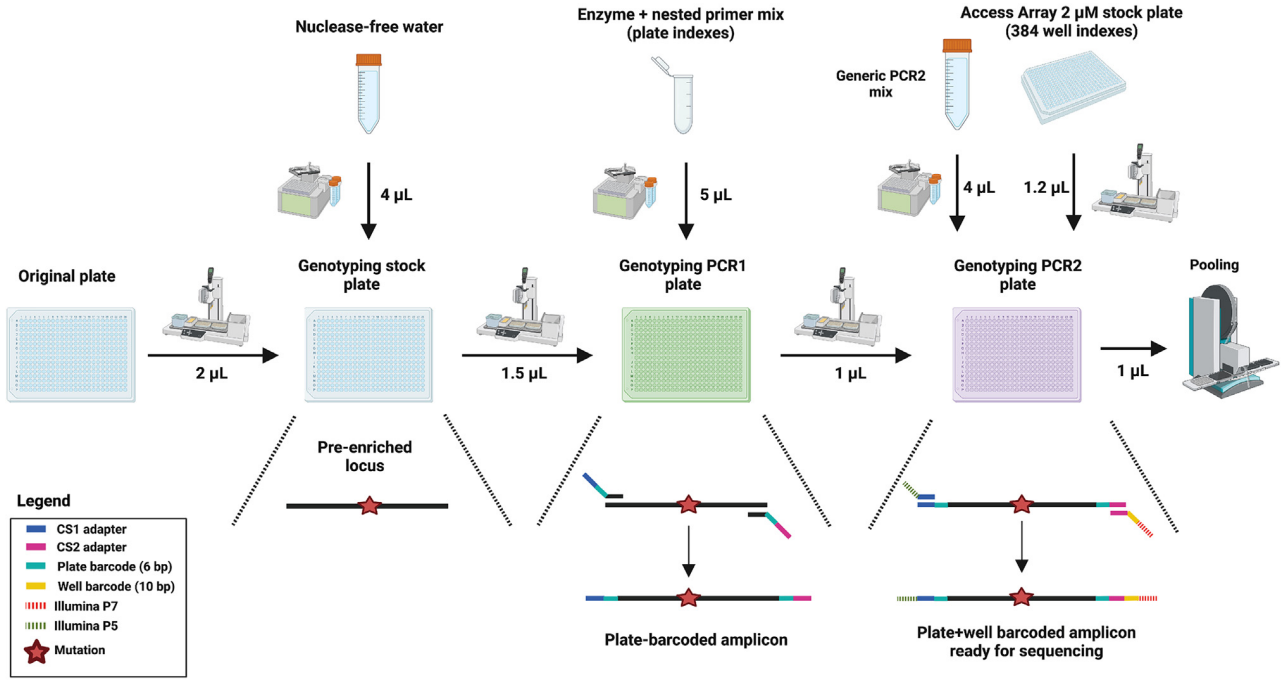


Figure 13. Single-cell genotyping procedures

Liquid handling platforms used at individual steps, as well as volumes transferred, are indicated next to the arrows. A schematic overview of the amplification and barcoding strategy at each step is shown below.

Note: Different samples and tissues display variable library complexity. While 200,000 reads/cell might be suitable for some tissue types, this may be either excessive or insufficient for others. PCR duplication values are usually a good metric of whether you sequenced the libraries to saturation.

Part 5: Single-cell genotyping library preparation

Timing: 2 days for 5–20 genotyping plates; 3 days for 20–40 genotyping plates

In this section, we describe the procedures for generating Illumina-compatible libraries for single-cell genotyping, starting from the genotyping stock plates prepared in part 3, point 38. For each cell, non-barcoded genotyping amplicons were generated in part 3, point 36 by target-specific genotyping primers. Now, genotyping amplicons need to be further enriched and fully barcoded, to be able to assign them to single cells. This is achieved via two additional PCR steps (Figure 13).

The first step (genotyping PCR1) uses primers that are nested within the original amplicon(s) and which contain universal CS1 (forward primer) and CS2 (reverse primer) adapters. Plate-specific 6 bp barcode sequences are incorporated into these primers, between the CS adaptors and the target-specific sequence, enabling libraries from different plates to be pooled together later. The second step (genotyping PCR2) uses primers that bind to the CS1/CS2 adapters (from the Access Array Barcode Library for Illumina Sequencers-384, Single Direction kit) and that attach Illumina-compatible sequencing adapters to genotyping amplicons. The primer that binds to the CS2 adapter also contains a 10 bp index, serving as cell barcode, which is well-specific. The combination of plate and cell barcode is unique for each cell across all plates, allowing all amplicons to be sequenced together. After genotyping PCR2, libraries are pooled, purified, and quantified for sequencing.

Note: The steps described in part 5 are performed using the MANTIS, INTEGRA VIAFLO, and the Mosquito HTS Nanolitre Liquid Handler (Figure 13). However, users might automate the protocol on other liquid handling platforms, such as the Biomek FxP Liquid Handling Platform (Beckman Coulter).[6]

61. Order barcoded genotyping PCR1 primers, validated in preparation two and preparation three (nested genotyping primers), for each genotyping locus of interest. Examples on how to design barcoded genotyping PCR1 primers[6] are provided in Table S5.
    a. Reconstitute barcoded genotyping PCR1 primers at 100 μM in TE buffer, in a designated pre-PCR clean area, ideally in a biosafety cabinet.
    b. Keep reconstituted primers at −20°C prior to use or proceed directly with point 62.

Note: The plate barcode is defined by the unique combination of the forward and the reverse 6 bp barcodes on the genotyping PCR1 primers. The number of different barcoded combinations of genotyping PCR1 primer pairs to order for each locus depends on the number of genotyping stock plates that you have. For example, if you sorted the sample into 10 plates, you need 10 different barcoded genotyping PCR1 primer pairs for each genotyping locus. For details, refer to Table S5.

⚠ CRITICAL: When ordering barcoded genotyping PCR1 primers, name each primer according to the locus and barcode, to avoid confusion. We also advise labelling the tube caps.

62. Prepare 770 μL of each genotyping PCR1 primer mix: add 8.58 μL of each 100 μM genotyping PCR1 primer to the respective mix and top up with nuclease-free water to 770 μL. A template for genotyping PCR1 primer mix preparation can be found in Table S6.

Note: At this point, the final concentration of each primer in the primer mix should be 1.11 μM.

Note: The total number of primer mixes depends on (a) the number of genotyping stock plates, and (b) the number of genotyped loci per cell. Each genotyping stock plate requires a different primer mix, due to different plate barcodes. Furthermore, in preparation three, point 21, you defined how many different genotyping reactions are required per single cell, which depends on the number of targeted loci. Usually, up to 4 different genotyping amplicons per cell can be combined in the same reaction. To provide examples:

a. If you have 5 genotyping plates and amplify 1 locus/cell, prepare 5 primer mixes;
b. if you have 5 genotyping plates and amplify 4 loci/cell, prepare 5 primer mixes;
c. if you have 5 genotyping plates and amplify 6 loci/cell, prepare 10 primer mixes;
d. if you have 10 genotyping plates and amplify 8 loci/cell, prepare 20 primer mixes.

Note: All primers used to amplify targets from a single initial genotyping stock plate need to have the same plate barcode, even if they are not all in the same genotyping PCR1 reaction (given that they barcode the same "original" plate). To summarize:

a. n° of different plate barcodes = n° of genotyping stock plates;
b. n° of primer mixes = n° of genotyping stock plates × n° of reactions per plate.

Note: We recommend preparing primer mixes in advance; this step can take several hours. We prepare primer mixes in a designated pre-PCR clean area to avoid contamination.

⚠ CRITICAL: If barcoded primers are added to the wrong mix, genotyping amplicon reads will be assigned to the wrong plate.

63. Prepare for genotyping PCR1:
    a. Clean all working surfaces with 70% ethanol.
    b. Thaw enough KAPA2G Robust HS Ready Mix and keep on ice.
    c. Vortex genotyping PCR1 primer mixes and keep on ice.
    d. Thaw genotyping stock plates prepared in part 3, point 38 and centrifuge them at 3,000 × *g* for 3 min at 21°C.
    e. Prepare sufficient fresh 384-well plates for the number of genotyping PCR1 reactions.
    f. Switch on the MANTIS liquid handling platform and the VIAFLO.
        i. Prepare a HV MANTIS chip used exclusively for PCR1.
        ii. Manually prime the chip with 500 µL of nuclease-free water (Methods video S1).

    Optional: Keep sorting plates, genotyping stock plates, genotyping PCR1, and genotyping PCR2 plates, all in different colors, to avoid confusion.

    ⚠ CRITICAL: Always centrifuge plates at 21°C and not at 4°C to avoid condensation on the plate cover, which may lead to cross-well contamination.

64. Set up the genotyping PCR1. Set up two genotyping PCR1 plates at a time as follows:

    Note: Here, we describe the procedure using the MANTIS and the VIAFLO liquid handling platforms.

    a. Prepare the enzyme-primer mix for one plate: mix 1300 µL of KAPA2G Robust HS Ready Mix with 700 µL of PCR1 primer mix (Table S6) and pipette mix 10 times.
    b. Aliquot 5 µL of enzyme-primer mix into each well of a fresh genotyping PCR1 plate, using the MANTIS.
    c. Place the aliquoted genotyping PCR1 plate and the genotyping stock plate on two plate holders of the VIAFLO, with the stage lever set to the central position (Figure 9).
        i. Transfer 1.5 µL of material from the genotyping stock plate into the genotyping PCR1 plate.
        ii. Mix 3 times using the same tips.
        iii. Discard the tips (Methods video S4).
    d. Seal the genotyping stock plate with an aluminum adhesive seal and the genotyping PCR1 plate with a plastic PCR adhesive seal. Centrifuge both plates in a plate spinner for 10 s at 500 × *g*. Snap freeze the genotyping stock plate on dry ice. Label the genotyping PCR1 plate appropriately and place it on wet ice.
    e. Change the non-filtered tip on the MANTIS HV chip.
        i. Wash with 600 µL of ethanol 70% and 600 µL of nuclease-free water.
        ii. Recover 6 µL 2–3 times.
        iii. Discard the tip.
        iv. Take the chip out of the holder and use tissue inserted in 70% ethanol to wipe both the bottom part of the chip and the place where the chip sits, to eliminate any trace of barcode.
        v. Dry the chip and the chip holder.
        vi. Return the chip to the holder and insert a fresh non-filtered tip into the chip. You are now ready to aliquot the next plate.
    f. Repeat points 64a–64e to prepare the second PCR1 plate. Use the same genotyping stock plate if you are preparing the second genotyping reaction for the same original sorting plate or switch genotyping stock plate if you are preparing the genotyping reaction for the next original sorting plate.

g. When two genotyping PCR1 plates are ready, place them in a thermal cycler with the lid heated at 105°C and run the following PCR1 program:

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	95°C	15 s	40 - n° cycles in first PCR (22–24 cycles in our case)
Annealing	60°C	20 s	
Extension	72°C	1 min	
Final extension	72°C	5 min	1
Hold	4°C	Hold	

h. Repeat points 64a–64g to prepare all genotyping PCR1 plates.

⚠ CRITICAL: Wash the MANTIS chip carefully between every genotyping PCR1 plate, to avoid contaminating the next genotyping PCR1 plate with plate barcodes from the previous plate.

Optional: If two people are working, we suggest that one person operates the MANTIS while the other operates the VIAFLO. This drastically speeds up the protocol. We also recommend processing all genotyping PCR1 plates before proceeding to generation of genotyping PCR2 plates.

Pause point: When genotyping PCR1 is done, you can proceed directly with genotyping PCR2, or you can snap freeze genotyping PCR1 plates and store them at −80°C for at least 3 months.

65. Prepare plates containing 1.2 µL of 2 µM barcoded PCR2 primers: these will be genotyping PCR2 plates.

Note: Here, we describe the procedure using the VIAFLO.

a. Prepare fresh 384-well plates (the number equivalent to that of genotyping PCR1 plates).
b. Thaw the Access Array 2 µM stock plate (preparation four) and centrifuge it at 3,000 × *g* for 3 min at 21°C.
c. Aliquot 1.2 µL of material from the Access Array 2 µM stock plate into each fresh 384-well plate, using the Repeat Dispense pipetting mode. As soon as four plates are aliquoted, cover them with adhesive aluminum seals, centrifuge in a plate spinner for 10 s at 500 × *g*, and place on wet ice.

⚠ CRITICAL: Aliquoted plates need to be covered and kept on ice, to avoid evaporation.

Pause point: You can either proceed with genotyping PCR2, or you can snap freeze plates prepared in point 65 on dry ice and store them at −80°C for at least 2 weeks.

66. Prepare enough genotyping PCR2 mix for the number of genotyping PCR2 plates you have, according to the table below. Keep the mix on ice.

Genotyping PCR2 mix

Reagent	1 reaction	1 plate (400 reactions)	Kit
FastStart Buffer 10× (labeled 3)	0.6 µL	240 µL	FastStart High Fidelity PCR System
$MgCl_2$ (labeled 4)	1.08 µL	432 µL	FastStart High Fidelity PCR System
DMSO (labeled 5)	0.3 µL	120 µL	FastStart High Fidelity PCR System

(*Continued on next page*)

Continued			
Reagent	1 reaction	1 plate (400 reactions)	Kit
Nucleotide mix (labeled 6)	0.12 μL	48 μL	FastStart High Fidelity PCR System
Enzyme (labeled 1)	0.06 μL	24 μL	FastStart High Fidelity PCR System
Nuclease-free water	1.84 μL	736 μL	N/A

Note: Thaw DMSO at 21°C – it tends to form aggregates if kept on ice.

67. Prepare for genotyping PCR2:
    a. Clean all working surfaces with 70% ethanol.
    b. Centrifuge genotyping PCR1 plates (point 64) and plates containing 1.2 μL of 2 μM barcoded PCR2 primers (point 65) at 3,000 × *g* for 2 min at 21°C.
    c. Switch on the MANTIS and the VIAFLO.
    d. Prepare a HV MANTIS chip used exclusively for PCR2. Manually prime the chip with 500 μL of nuclease-free water.
68. Set up two genotyping PCR2 plates at a time.

Note: Here, we describe the procedure using the MANTIS and the VIAFLO.

    a. Place the plate containing 1.2 μL of 2 μM barcoded PCR2 primers on a cold rack and on the MANTIS plate holder, and aliquot 4 μL of genotyping PCR2 mix into this plate. Hereafter, we refer to the aliquoted plate as to a genotyping PCR2 plate.
    b. Cover the plate with an aluminum or plastic adhesive seal and centrifuge in a plate spinner for 10 s at 500 × *g*.
    c. Place one genotyping PCR1 plate and the aliquoted genotyping PCR2 plate on two plate holders of the VIAFLO with the stage lever set to the central position (Figure 9).
        i. Transfer 1 μL of material from the genotyping PCR1 plate to the genotyping PCR2 plate.
        ii. Mix 3 times using the same tips.
        iii. Discard the tips.
    d. Seal the genotyping PCR1 plate with an aluminum adhesive seal and the genotyping PCR2 plate with a plastic PCR adhesive seal. Centrifuge both plates in a plate spinner for 10 s at 500 × *g*, snap freeze the genotyping PCR1 plate, and place the genotyping PCR2 plate on wet ice.
    e. Repeat points 68a–68d to prepare a second genotyping PCR2 plate, using a second genotyping PCR1 plate. There is no need to wash the MANTIS chip in between plates.
    f. When two genotyping PCR2 plates are done, place them in a thermal cycler with the lid heated at 105°C and run the following PCR2 program:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	10 min	1
Denaturation	95°C	15 s	10 cycles
Annealing	60°C	30 s	
Extension	72°C	1 min	
Final extension	72°C	5 min	1
Hold	4°C	Hold	

    g. Repeat points 68a–68f to prepare all genotyping PCR2 plates.

Optional: If two people are working, we suggest that one person operates the MANTIS while the other operates the VIAFLO. This setup will drastically speed up the execution of the

protocol. We also recommend preparing all genotyping PCR2 plates before proceeding to the next points.

Note: Wipe the chip with paper inserted in 70% ethanol between every 4 genotyping PCR2 plates, to clean potential splashes of barcoded PCR2 primers.

Pause point: When genotyping PCR2 is done, you can proceed, or you can snap freeze genotyping PCR2 plates on dry ice and store them at −80°C for at least 3 months. We usually process ~40 genotyping PCR2 plates per day.

Note: Each well of each genotyping PCR2 plate is now fully barcoded with a combination of well and plate barcode. However, we pool and purify each genotyping PCR2 plate separately, for QC purposes.

69. Pool the wells from genotyping PCR2 plates (1 μL from each well).

Note: Here, we describe the procedure using the Mosquito HTS Nanolitre Liquid Handler.

a. Centrifuge all genotyping PCR2 plates in a plate spinner for 10 s at 500 × *g* and keep them on ice.
b. Prepare a fresh 384-well plate, used for genotyping PCR2 plate pooling, and label it as "PCR2 pooling plate". The goal is to pool genotyping PCR2 plates into alternate columns of the PCR2 pooling plate, in which each column will contain material from one full genotyping PCR2 plate (Figure 14A).
c. Switch on the Mosquito and set up a program as in Figure 14B: this will pool an entire PCR2 plate into column 1 of the PCR2 pooling plate, taking 1 μL from each well of genotyping PCR2 plate.
d. Insert one genotyping PCR2 plate and the PCR2 pooling plate into magnetic racks for the Mosquito (Figure 14C).
e. Place the genotyping PCR2 plate at position 1 and the PCR2 pooling plate at position 2 of the Mosquito (Figure 14C).
f. Run the program.
g. When the program is done, cover the genotyping PCR2 plate with an aluminum adhesive seal and snap freeze.
h. Repeat points 69d–69g for the second genotyping PCR2 plate, this time pooling into column 3 of the PCR2 pooling plate (Figure 14A). The third genotyping PCR2 plate will be pooled into column 5, etc.

Note: Given that you pool into every second column of the PCR2 pooling plate, you can pool up to 12 genotyping PCR2 plates into one pooling plate.

Note: Switch Mosquito tips between each genotyping PCR2 plate to maintain pipetting accuracy.

Optional: Genotyping PCR2 plate pooling can also be performed with a multichannel pipette or with an alternative liquid handling platform.

70. When the PCR2 pooling plate is ready, pool all the wells of a single column into a fresh DNA LoBind 1.5 mL microcentrifuge tube (or equivalent), labeled according to the genotyping PCR2 plate, using a P200. The total number of tubes equals the number of genotyping PCR2 plates.

Optional: If you have many columns to pool, we suggest using a 12 × 200 μL multichannel pipette and transferring the material into a V-bottom 96-well plate. Keep track of which pool is in which well.

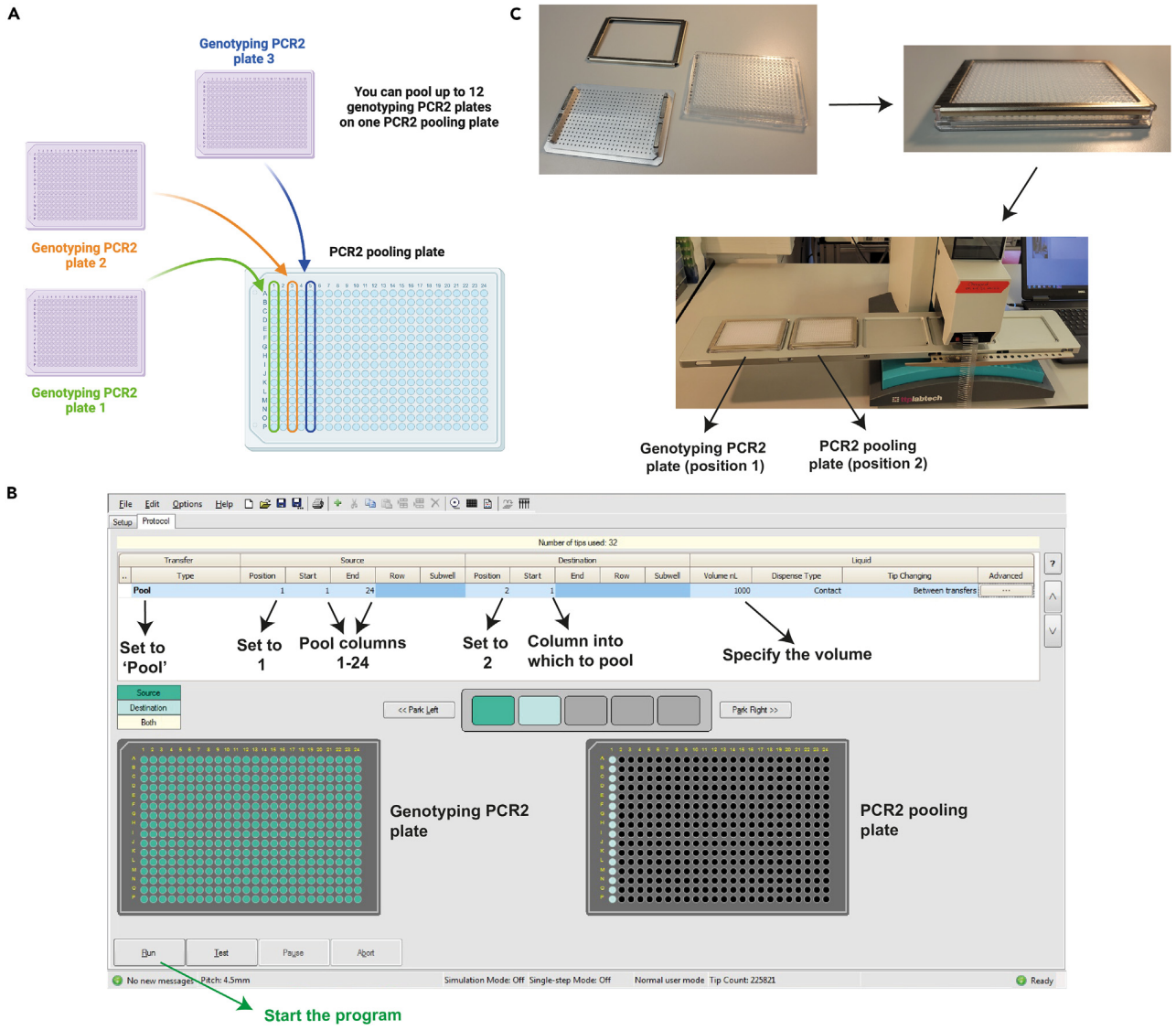


Figure 14. Mosquito setup for genotyping PCR2 plate pooling

(A) Schematic overview of the procedure for pooling genotyping PCR2 plates.

(B) Mosquito program for genotyping PCR2 plate pooling.

(C) Setup of plates on the Mosquito.

71. Prepare for bead purification of the pools using V-bottom 96-well plates:
    a. Incubate AMPure XP beads for 30 min at 21°C. After 30 min, vortex thoroughly and pour into a clean reservoir, labeled "beads".
    b. Prepare 80% (vol/vol) ethanol by mixing 40 mL of ethanol 100% and 10 mL of nuclease-free water. Pour 80% ethanol into a clean reservoir, labeled "EtOH".
    c. Pour nuclease-free water into a second clean reservoir, labeled "H_2O".
    d. Transfer 100 µL of each pool to a well of a fresh V-bottom 96-well plate.

Optional: You can also bead-purify in microcentrifuge tubes or in PCR strips, but for higher numbers of libraries, using an 8 × 200 µL multichannel pipette and the V-shaped 96-well plates is more practical.

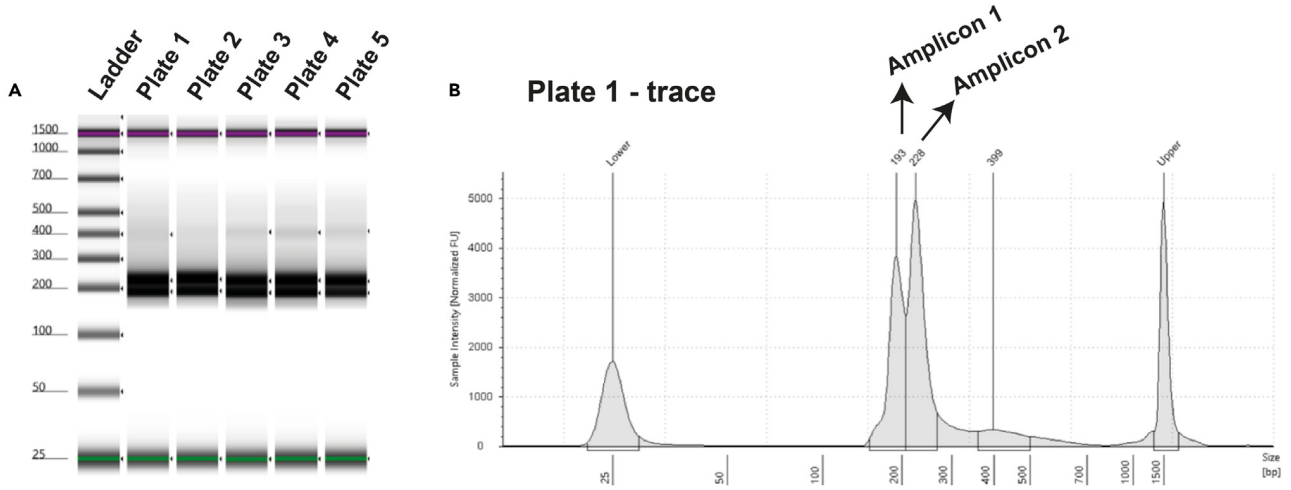


Figure 15. Quality control of genotyping libraries

(A) TapeStation bands showing genotyping amplicons for each plate (382 pooled cells). Amplicons of expected size should be observed.

(B) A TapeStation trace showing intensity of signal as function of size in bp.

72. Add 120 µL of beads (1.2× beads:sample ratio) to each pool and mix 20–30 times. Avoid bubbles.
    a. Incubate at 21°C for at least 5 min – this will allow DNA fragments of interest to bind to the beads.
73. Place the plate on a magnetic stand for plates. Beads will accumulate towards the magnet. Wait for the liquid to become completely clear of beads (minimum 3 min).
74. When the supernatant is completely clear, carefully remove and discard all supernatant (~220 µL), taking care not to aspirate the beads.
75. Proceed immediately with 80% ethanol washing:
    a. Dropwise, add 200 µL of 80% ethanol to each pool, not disturbing the beads.
    b. Incubate for 30 s.
    c. Remove and discard all 80% ethanol, not disturbing the beads.
    d. Repeat points 75a–75c 1–2 more times.
    e. After the final wash, carefully remove traces of ethanol using 20 µL tips.
76. Let the beads air-dry for a few minutes (usually 4–5 min at this step). The beads are dry enough when the surface of the pellet changes from shiny to matt.
77. Remove the plate from the magnet. Resuspend beads in 50 µL of nuclease-free water. Mix thoroughly 20–30 times to completely resuspend the beads and incubate at 21°C for 3 min.
78. Place the plate back on the magnet and wait for 2–3 min for the supernatant to be completely devoid of beads. Libraries are now in the supernatant.
79. Transfer 48–50 µL of clear supernatant to fresh DNA LoBind 1.5 mL microcentrifuge tubes (or equivalent), labeled accordingly. Take care not to transfer any beads.
80. Measure the concentration of each pool using the Qubit dsDNA HS Assay Kit. Use 1 µL of each library. Take note of the pool concentration [ng/µL].
81. For each pool, check genotyping library size distribution using a microcapillary assay such as Agilent Bioanalyzer or TapeStation with a D1000 Kit. Check that each pool contains peaks at the expected size and record the average library size by choosing the region between 50 and 900 bp. Representative good quality libraries can be found in Figure 15.

Note: Barcodes and adapter sequences add a total of 115 extra bp to the genotyping amplicon.

Pause point: You can freeze the pools and keep them at −20°C for 6 months.

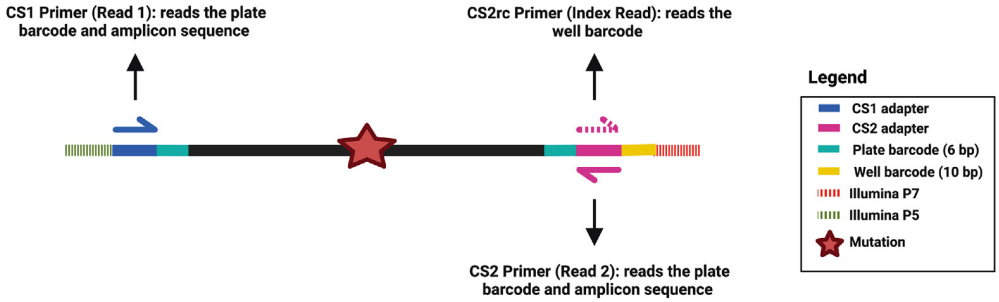


Figure 16. Sequencing strategy for genotyping libraries

CS2rc is the reverse complement of the CS2 sequence. Note that, if sequencing on the NextSeq, a longer version of the CS1 primer (LCS1) is used instead of the CS1 primer.

82. Combine equimolar ratios of each pool to a fresh DNA LoBind 1.5 mL microcentrifuge tube (or equivalent), for pools that will be sequenced on the same sequencing run. When pooling, consider the average library size, the concentration [ng/µL], and the number of amplicons per pool. Use Table S7 as template.

Optional: At this stage, you may perform a double-size AMPure XP bead selection on the final pool, by using a 0.6×, followed by a 1.2× beads:sample ratio. This step can help remove or reduce the high-bp peak which is likely a trace of the scATAC-seq library. However, do not perform this step if you expect a genotyping amplicon to be of that size!

83. Measure the concentration [ng/µL] of the final pool with Qubit and the average library size with Agilent TapeStation D1000.
    a. Dilute the library to 4 nM for sequencing (Table S7).
84. Sequence the library on either a MiSeq platform, using custom sequencing CS1 (read 1), CS2 (read 2), and CS2rc (index read) primers, or on a NextSeq platform, using custom sequencing LCS1 (read 1), CS2 (read 2), and CS2rc (index read) primers (Figure 16). Single-use primer aliquots were prepared in preparation four.

Note: The choice of kit depends on the desired sequencing depth: we usually aim for 2,000 reads per cell per amplicon.

    a. For the MiSeq:
        i. Pierce ports 12, 13 and 14 containing Illumina sequencing primers with separate 1 mL tips. Aspirate the entire volumes into 3 tubes, labeled I12, I13, and I14.
        ii. Mix 10 µL of 100 µM CS1 with 10 µL of 100 µM CS2 to obtain 20 µL of CS1/CS2 50 µM (we refer to CS1/CS2 50 µM as to "FL1").
        iii. Mix 7 µL of FL1 with 200 µL of I12 and with 493 µL of HT1 buffer. Vortex and centrifuge briefly. Transfer the entire volume into port 18.
        iv. Mix 3.5 µL of CS2rc 100 µM with 200 µL of I13 and with 496.5 µL of HT1 buffer. Vortex and centrifuge briefly. Transfer the entire volume into port 19.
        v. Mix 7 µL of FL1 with 200 µL of I14 and with 493 µL of HT1 buffer. Vortex and centrifuge briefly. Transfer the entire volume into port 20.
        vi. Denature and dilute the 4 nM library from point 83a following the Illumina guidelines. Spike in 5%–10% of PhiX.
    b. For the NextSeq:
        i. Pierce ports 20, 21 and 22 containing Illumina sequencing primers with separate 1 mL tips. Aspirate the entire volumes into 3 tubes, labeled I20, I21, and I22.
        ii. Mix 6 µL of LCS1 100 µM primer with 800 µL of I20 and with 1194 µL of HT1. Vortex and centrifuge briefly. Transfer the entire volume into port 7.

iii. Mix 6 μL of CS2 100 μM primer with 800 μL of I21 and with 1194 μL of HT1. Vortex and centrifuge briefly. Transfer the entire volume into port 8.
iv. Mix 6 μL of CS2rc 100 μM primer with 1994 μL of HT1. Vortex and centrifuge briefly. Transfer the entire volume into port 9.

Note: I22 is not required for sequencing PhiX and does not need to be added to the custom CS2rc index primer.

v. Denature and dilute the 4 nM library from point 83a following the Illumina guidelines. Spike in 5%–10% of PhiX.

Note: For primer sequences, refer to the key resources table. Use the following sequencing configuration: 76 cycles (read 1) + 10 cycles (index read) + 76 cycles (read 2).

⚠ CRITICAL: Make sure to perform enough cycles to read the mutant locus at least with read 1 or read 2.

⚠ CRITICAL: Make use of a PhiX spike-in (5%–10%), given the low genotyping library complexity! If PhiX is not used, the sequencing run is likely to fail.

Note: CS1/LCS1 and CS2 primers will bind to CS1 and CS2 adaptors inserted during PCR1, allowing the amplicon and plate barcodes to be sequenced. The CS2rc primer is the reverse complement of the CS2 adaptor allowing the 10 bp cell index inserted during PCR2 to be sequenced.

Optional: Given that nested genotyping amplicons are usually short (100–200 bp for most amplicons), we perform 75 bp paired-end sequencing, to avoid read 1 or read 2 sequencing the CS adapters. However, in case of longer amplicons, you can use a different configuration – the key is that either read 1 or read 2 successfully cover the mutant locus of interest.

EXPECTED OUTCOMES

Single-cell ATAC-seq library generation: after part 4, point 57, you should expect to generate scATAC-seq library traces comparable to those in Figure 2. A sign of a good quality library is the presence of multiple peaks displaying the characteristic nucleosomal pattern in a wavy shape. Peak heights may be variable and depend on sample and tissue type. After bead purification, when libraries are eluted in 22 μL, we usually obtain ~10–30 ng/μL of DNA. Note that these numbers apply to human bone marrow samples processed with 16 cycles of PCR. We observed higher yields when processing cell lines. In our experience, library shape is more informative of successful library generation than the concentration. If sequencing on the NextSeq, you should expect at least 85% of bases > Q30.

After running the pre-processing pipeline, you should expect ~90% or more wells passing QC filters. A successful human bone marrow cell sequenced at the depth of ~200,000 reads usually displays: > 95% mapping rate; 40,000–150,000 unique nuclear fragments/cell; 0.5%–10% of reads mapping to the mitochondrial genome; 50%–80% of mapped reads in called peaks; TSS enrichment > 8 as computed by ArchR.[7] A duplication level of > 70% indicates that libraries are sequenced close to saturation.

Single-cell genotyping library generation: after part 5, point 80, each genotyping library should contain 5–100 ng/μL of DNA. Note that the yield is likely to depend on the number of loci amplified in each reaction. After part 5, point 81, for each genotyping library, you should observe bands of the expected size, as shown in Figure 15. We frequently observed an additional low-intensity band

between ~400 bp and 500 bp – this is usually not problematic. In the case of multiple amplicons in the same mix, it is common to observe that some bands are more intense than others, due to differences in amplification efficiency for different loci. This is usually not a point of concern, and lower intensity bands still allow for efficient genotyping. If sequencing on the NextSeq, you should expect at least 85% of bases > Q30. In a good quality run, 95%–100% of cells that pass scATAC-seq QC should also be successfully genotyped for the given amplicon.

QUANTIFICATION AND STATISTICAL ANALYSIS

1. Pre-processing of scATAC-seq data:
   a. For scATAC-seq library demultiplexing, prepare a sample sheet containing i7 and i5 barcode information for each single cell, according to Illumina guidelines. Troubleshooting 12.
   b. Demultiplex raw sequencing data with bcl2fastq, which generates two FASTQ files per cell (R1 and R2).
   c. These files are used as input for the previously published pre-processing pipeline[2] (https://github.com/dbrg77/scATAC_snakemake).
      i. We use a slightly modified version of this pipeline, in which we do not remove duplicate reads from the merged BAM file (we still remove them from BAM files corresponding to single cells).
      ii. Note that you may use other pre-processing strategies, which may involve a different read mapping tool, for example.
   d. The pipeline generates a fragment file, containing coordinates and cell barcodes of all nuclear fragments.
   e. This file is used as input for downstream analysis pipelines, like ArchR.[7] Alternatively, you may use Signac[11] or other packages for downstream analysis.
      i. You can find our scripts for downstream analysis using ArchR and Signac at https://github.com/sventurkalj/GTAC/tree/v1.0.0.
      ii. At this step, filtering of low-quality cells is necessary prior to downstream analysis. Filtering thresholds will depend on sample quality and are usually estimated on per-sample basis. In our experiments, we exclude inferred doublets (usually 1%–5% of total cells) and include cells displaying between 10,000 and 120,000 unique nuclear fragments, and TSS enrichment > 8. These parameters are rather stringent and can be relaxed based on data distribution and sequencing depth. However, we advise excluding at least doublets and cells with < 1,000 unique nuclear fragments and TSS enrichment < 4.
2. For single-cell genotyping pre-processing, we use a previously published pipeline used for TARGET-seq[3,6] (https://github.com/albarmeira/TARGET-seq):
   a. The GenoDemux_Fastq.sh script will perform demultiplexing. As input, you need:
      i. The generic "SampleSheet_byWell_TARGET.csv" file containing the 384 well barcodes inserted during genotyping PCR2.
      ii. The metadata file containing information about plate and well barcode for each single cell. Here, it is important to keep cell IDs (names) consistent with the ones used for scATAC demultiplexing, to be able to later link genotypes to accessibility profiles of single cells. Troubleshooting 12.
   b. Once two FASTQ files have been generated per cell (R1 and R2), proceed with the SCgenotype.pl script, following the instructions. Note that, in GTAC, we do not use mRNA primers, hence you can leave the mRNA primer files blank. The pre-processing pipeline will generate a summary file of nucleotide counts per single cell, for each locus of interest.
   c. These summary files are used as input for our custom genotyping analysis scripts (https://github.com/sventurkalj/GTAC/tree/v1.0.0):
      i. We first compute the total locus coverage per cell and exclude cells with less than 30 reads across the locus.

ii. We then compute single cell variant allele frequencies (scVAF) and mutant read numbers across control WT cells to set thresholds for mutation calling based on these values.[1] This will result in a cell being called mutant or WT at the locus.
iii. If multiple loci are screened, we repeat this procedure for all loci. Next, we integrate mutation calls across all loci for each single cell and generate a matrix in order to run infSCITE, which identifies the most likely clonal hierarchy. Note that alternative approaches for tree inference can be used.
iv. The second custom script then assigns each cell to a clone, based on the combination of mutations in that cell. Here, the conditions set for clonal assignment are driven by the infSCITE result. For a small proportion of cells, a clone cannot be confidently assigned due to failed genotyping on multiple loci – we exclude these cells from downstream analysis (for details refer to Turkalj, Jakobsen et al., 2023[1]).

3. Note that, while the pre-processing pipeline can readily detect SNVs, indels may be challenging to count accurately.
   a. In those cases, to get WT and mutant counts from single cells, you can apply the grep function in Unix directly to the FASTQ files, counting the number of WT and indel sequences observed in each file.
      i. If the indel is covered both by R1 and R2, grep the WT and mutant sequences from both FASTQ files. Note that, for R2, you need to provide the reverse complement sequences.
      ii. In those cases, the total WT and mutant coverage is given by the sum of reads from the R1 and R2 files.
   b. For specific indels, like FLT3-ITD, specialized pipelines[12] can be adapted.
4. Finally, clone information for each single cell is merged with its scATAC-seq metadata via cell IDs. This will allow clone-specific analysis of chromatin accessibility in downstream steps.
   a. Our custom pipelines (https://github.com/sventurkalj/GTAC/tree/v1.0.0) provide examples of code used for downstream clone-specific analysis of chromatin accessibility.

LIMITATIONS

At present, GTAC has several limitations. First, in terms of cell numbers, GTAC provides lower throughput compared to droplet-based strategies. At best, one tagmentation reaction on 50,000–100,000 cells can yield up to ten FACS sorted 384-well plates. If starting cell numbers are high, you can decide to use multiple tagmentation reactions, to sort more plates. Even if almost completely automated, the method requires around 5 days to fully process ten 384-well plates. However, the time required increases with increasing numbers of genotyped loci, given that the number of genotyping PCR1 and PCR2 reactions per cell increases. Altogether, data from 10,000–12,000 cells can be generated in a few weeks. However, most sorted nuclei usually pass QC filters, both for genotyping and scATAC-seq libraries, partially compensating for lower throughput. Other methods achieving higher throughput[13] are available but currently suffer from lower genotyping sensitivity and scATAC-seq library content.

Second, GTAC relies on targeted analysis of mutations previously detected in the same samples by either targeted panels or whole exome sequencing of cell populations, meaning that a portion of the initial sample needs to be used for driver mutation discovery prior to GTAC.

Third, for genetically complex tumors, in which hundreds of mutations are present within the same sample, it would be necessary to limit GTAC to a subset of mutations. In our study,[1] we successfully genotyped 7 genomic loci simultaneously in a primary sample. Notably, the scATAC-seq library shape was unaltered even when 10, 15, or 20 loci were amplified in parallel (Figure 17). However, applying GTAC to large numbers of loci would require further validation.

Finally, GTAC does not enable the capture of RNA from the same cells, limiting our understanding of how chromatin variation, associated with different mutations, alters gene expression. If multiple vials

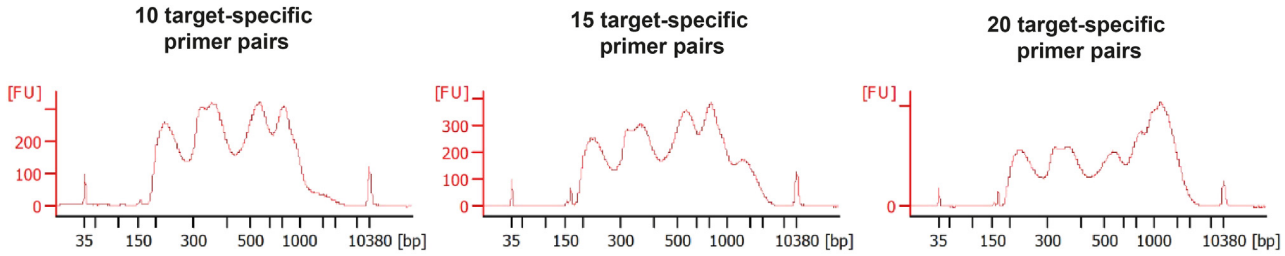


Figure 17. Bioanalyzer traces showing ATAC libraries when 10, 15, or 20 genomic loci were co-amplified
Human CD34+ bone marrow cells were used. Each trace derives from 96 pooled single cells.

from the same sample are available, methods integrating single-cell genotyping or clonal inference with transcriptomic readouts can be applied,[3,14–17] and scRNA and scATAC datasets can be integrated computationally. However, future methods integrating single-cell genotyping, chromatin accessibility, and transcriptome readouts from the same cells would be beneficial to tackle these limitations.

TROUBLESHOOTING

Problem 1

scATAC-seq library yield is very low and insufficient for sequencing (part 4, points 56–57; Figure 18A).

Potential solution

- The number of PCR cycles (part 3, point 36) is insufficient. Increase the number of PCR cycles based on the results obtained in preparation one. If your scATAC-seq library yield is still low, if possible, perform a qPCR specifically on the problematic sample, to define the optimal number of cycles for that specific sample, using a few columns of one sorted plate.
- The sorter is not properly calibrated (part 2, point 27), resulting in excessive empty wells. Calibrate carefully until the droplets are collected at the center of the well. You can also estimate the sorting accuracy by simulating a single-cell genotyping test on your sorted plate, as in preparation three (use 16–32 cells from few random columns) and checking how many wells show a genotyping band.

Problem 2

Bioanalyzer trace shows a large low-bp peak. ATAC trace is either of low intensity or absent (part 4, point 57; Figures 18B and 18C).

Potential solution

- It is likely that some of your pre-amplification genotyping primers bind with high affinity to the i7 and/or i5 primers, thus interfering with scATAC-seq library amplification. Test all pre-amplification genotyping primers individually as described in preparation three, to identify the problematic pair. Avoid primers for which the predicted secondary structure of the heterodimer formed with the i7 or i5 primer displays $\Delta G < -9$ kcal/mol (IDT Oligo Analyzer; preparation two). Make sure you are not using excessive primer concentration.

Problem 3

Bioanalyzer shows no ATAC trace at all (part 4, point 57; Figure 18D).

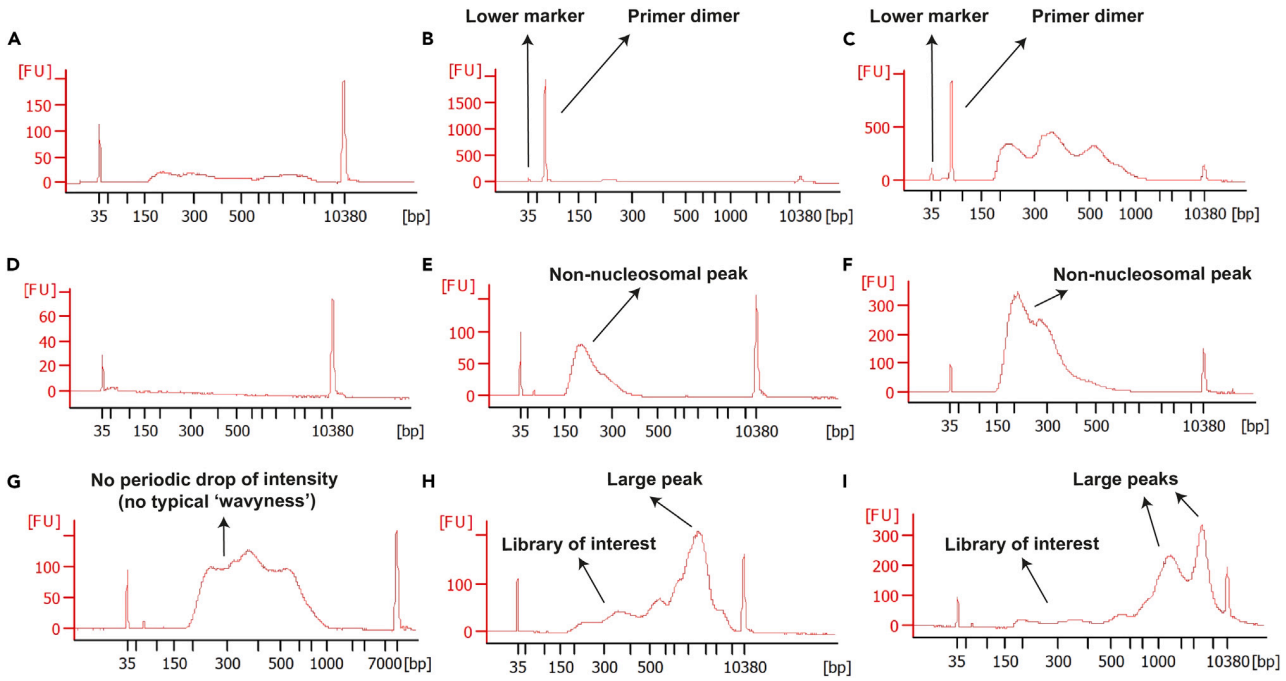


Figure 18. Representative suboptimal ATAC traces

(A)–(I) Bioanalyzer traces displaying failed or suboptimal ATAC library amplification, which requires troubleshooting.

Potential solution

- Make sure to use a Tn5 enzyme which is not expired. Make sure to add enough Tn5 enzyme to the tagmentation reaction (part 2, points 24g and 25).
- Make sure to have sufficient $MgCl_2$ in the tagmentation mix.
- Make sure to include the 72°C gap-filling step in the PCR (part 3, point 36).
- Make sure that the i7 and i5 primer sequences that you are using are correct.
- Make sure the sorter is properly calibrated (refer to potential solution of problem 1).

Problem 4

ATAC trace displays only a prominent non-nucleosomal peak (Figure 18E) or only a non-nucleosomal peak and a lower, mono-nucleosomal peak (part 4, point 57; Figure 18F). This is likely to result in low library complexity.

Potential solution

- Increase the elongation time of the PCR (part 3, point 36) up to 1 min. Some samples show a preponderance of non-nucleosomal compared to nucleosomal fragments and may require PCR optimization. Having only the non-nucleosomal peak will reduce library content.
- This might also indicate over-tagmentation. Make sure to add 2× tagmentation stop buffer after tagmentation (part 2, point 26). Make sure that the EDTA concentration in the 2× tagmentation stop buffer is appropriate.

Problem 5

ATAC trace does not display the typical wavy shape: periodic drops of intensity are not prominent. (part 4, point 57; Figure 18G). This is likely to result in low library complexity. Sequencing may show a high percentage of mitochondrial reads. Signal-to-noise ratio (FRiP) may be low.

Potential solution

- It is likely your detergent lysis conditions are suboptimal, leading to excessive lysis and disruption of nucleosomes prior to tagmentation. Remake all detergent solutions, taking care not to add excessive detergents.
- Make sure your thermal mixer is set at 37°C during tagmentation.
- When resuspending cells and nuclei, pipette gently not to disrupt the structure of cells or nuclei.

Problem 6

ATAC trace displays a very intense high-bp peak (part 4, point 57; Figures 18H and 18I).

Potential solution

- Repeat the double-side size selection with AMPure XP beads on your libraries (part 4, points 46–55), using first a 0.5×, followed by a 1.2× beads:sample ratio.
- If this is recurrent even after double-side size selection, consider reducing elongation time of the PCR (part 3, point 36) by 5–10 s

Problem 7

A very low number of nuclei in the tube prior to single-nucleus sorting (part 2, point 28).

Potential solution

- Take extreme care when aspirating the pellets during the Omni-ATAC procedure (part 2, point 24). Do not disturb pellets. Make sure to use a fixed-angle centrifuge. If you can't see the nuclear pellet, leave a few µL of supernatant inside the tube.

Problem 8

Many non-lysed cells prior to single-nucleus sorting (part 2, point 28).

Potential solution

- Make sure to use the correct concentration of all detergents, and to lyse the cells for 3 min, when performing the Omni-ATAC procedure (part 2, point 24).

Problem 9

Sequencing shows many empty wells (quantification and statistical analysis).

Potential solution

- Make sure the sorter is properly calibrated (refer to potential solution of problem 1).
- Make sure your DAPI-positive gate is stringent enough (part 2, point 29c; Figure 11).

Problem 10

Sequencing shows excessive mitochondrial read percentage (quantification and statistical analysis).

Potential solution

- Collect all, or almost all, supernatant when aspirating liquid during the Omni-ATAC procedure (part 2, point 24h). Mitochondrial DNA likely resides mostly in the supernatant.

CellPress
OPEN ACCESS

Problem 11

Many cells show a successful scATAC-seq library but poor or failed genotyping at one or more loci (quantification and statistical analysis).

Potential solution

- This indicates that, while the sort and scATAC-seq library construction went well, there is a problem in genotyping library construction. Revisit your primer testing results (preparation two and preparation three) and evaluate genotyping bands for the problematic loci. If bands are faint, design new primer pairs and select those that generate bright bands.
- Increase pre-amplification genotyping primer concentration up to 400 nM and test whether bands are brighter (preparation three, point 15).
- Change the organization of your genotyping PCR1 reactions: if there is one problematic locus, amplify that locus separately from the remaining loci, in a separate genotyping PCR1 reaction.

Problem 12

Empty control wells show many sequencing reads, and/or the control WT sample shows many mutant reads, on one or several plates (quantification and statistical analysis).

Potential solution

- This likely indicates cross-well contamination in those plates. Data from these plates should not be used and cannot be trusted. Always centrifuge plates prior to unsealing. When unsealing, take care not to generate any liquid splashes (keep the plate firmly against the bench). Always centrifuge plates at 21°C to avoid condensation of the seals.
- Alternatively, you might be assigning sequencing reads to the wrong wells (due to a mistake in the sample sheet or in the genotyping metadata; quantification and statistical analysis).

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Paresh Vyas (paresh.vyas@imm.ox.ac.uk).

Materials availability

The list of previously validated target-specific genotyping primers[1] can be found in Table S1. The list of barcoded i7[3] and i5 oligo sequences used for scATAC-seq library amplification can be found in Table S2.

Data and code availability

The code for GTAC analysis can be accessed at GitHub: https://github.com/sventurkalj/GTAC/tree/v1.0.0. The code is also provided in an open access disposition at Zenodo: https://doi.org/10.5281/zenodo.7817225. Python and Perl scripts used for scATAC-seq or single-cell genotyping data pre-processing were published previously[2,3] and are available at the respective GitHub pages. DOIs are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2023.102641.

ACKNOWLEDGMENTS

P.V. acknowledges funding from the Medical Research Council Molecular Haematology Unit Programme Grant (MC_UU_00029/8), Blood Cancer UK Programme Continuity Grant 13008, NIHR Senior Fellowship, and the Oxford BRC Haematology Theme. S.T. is supported by a Scatcherd

European Scholarship in partnership with the Medical Research Council/Radcliffe Department of Medicine and the Clarendon Fund. N.A.J. is supported by a Medical Research Council and Leukaemia UK Clinical Research Training Fellowship (MR/R002258/1). A.G. is supported by DPhil studentship funding from the Lady Tata Memorial Trust. F.A.R. is supported by the Oxford – Sir David Weatherall Scholarship from Green Templeton College and the WIMM Prize Studentship from the MRC Weatherall Institute of Molecular Medicine. The authors thank Dr. Felix Dingler, Dr. Assunta Adamo, and Mr. Ramy Slama for insightful discussions. The authors also acknowledge the MRC WIMM Flow Cytometry Facility and Single Cell Facility. Figures, or part of Figures 1, 3, 5, 7, 13, 14, and 16, as well as the Graphical abstract, were created using BioRender.

AUTHOR CONTRIBUTIONS

Conceptualization, S.T., N.A.J., and P.V.; methodology, S.T., N.A.J., and A.G.; investigation, S.T., N.A.J., and A.G.; formal analysis, S.T. and N.A.J.; visualization, S.T. and F.A.R.; supervision, P.V.; writing – original draft, S.T., N.A.J., A.G., and F.A.R.; writing – review and editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Turkalj, S., Jakobsen, N.A., Groom, A., Metzner, M., Riva, S.G., Gür, E.R., Usukhbayar, B., Salazar, M.A., Hentges, L.D., Mickute, G., et al. (2023). GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells. Cell Stem Cell *30*, 722–740.e11. https://doi.org/10.1016/j.stem.2023.04.012.

2. Chen, X., Miragaia, R.J., Natarajan, K.N., and Teichmann, S.A. (2018). A rapid and robust method for single cell chromatin accessibility profiling. Nat. Commun. *9*, 5345. https://doi.org/10.1038/s41467-018-07771-0.

3. Rodriguez-Meira, A., Buck, G., Clark, S.A., Povinelli, B.J., Alcolea, V., Louka, E., McGowan, S., Hamblin, A., Sousos, N., Barkas, N., et al. (2019). Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. Mol. Cell *73*, 1292–1305.e8. https://doi.org/10.1016/j.molcel.2019.01.009.

4. Xu, W., Wen, Y., Liang, Y., Xu, Q., Wang, X., Jin, W., and Chen, X. (2021). A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility. Nat. Protoc. *16*, 4084–4107. https://doi.org/10.1038/s41596-021-00583-5.

5. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinf. *13*, 134. https://doi.org/10.1186/1471-2105-13-134.

6. Rodriguez-Meira, A., O'Sullivan, J., Rahman, H., and Mead, A.J. (2020). TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. STAR Protoc. *1*, 100125. https://doi.org/10.1016/j.xpro.2020.100125.

7. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.

8. Jahn, K., Kuipers, J., and Beerenwinkel, N. (2016). Tree inference for single-cell data. Genome Biol. *17*, 86. https://doi.org/10.1186/s13059-016-0936-x.

9. Xu, W., Yang, W., Zhang, Y., Chen, Y., Hong, N., Zhang, Q., Wang, X., Hu, Y., Song, K., Jin, W., and Chen, X. (2022). ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.16.476488.

10. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962. https://doi.org/10.1038/nmeth.4396.

11. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.

12. Blätte, T.J., Schmalbrock, L.K., Skambraks, S., Lux, S., Cocciardi, S., Dolnik, A., Döhner, H., Döhner, K., and Bullinger, L. (2019). getITD for FLT3-ITD-based MRD monitoring in AML. Leukemia *33*, 2535–2539. https://doi.org/10.1038/s41375-019-0483-z.

13. Myers, R.M., Izzo, F., Kottapalli, S., Prieto, T., Dunbar, A., Bowman, R.L., Mimitou, E.P., Stahl, M., El Ghaity-Beckley, S., Arandela, J., et al. (2022). Integrated Single-Cell Genotyping and Chromatin Accessibility Charts JAK2 V617F Human Hematopoietic Differentiation. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.11.491515.

14. Nam, A.S., Kim, K.T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R.M., Abu-Zeinah, G., Brand, R., Omans, N.D., et al. (2019). Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature *571*, 355–360. https://doi.org/10.1038/s41586-019-1367-0.

15. Velten, L., Story, B.A., Hernández-Malmierca, P., Raffel, S., Leonce, D.R., Milbank, J., Paulsen, M., Demir, A., Szu-Tu, C., Frömel, R., et al. (2021). Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. Nat. Commun. *12*, 1366. https://doi.org/10.1038/s41467-021-21650-1.

16. Beneyto-Calabuig, S., Merbach, A.K., Kniffka, J.-A., Antes, M., Szu-Tu, C., Rohde, C., Waclawiczek, A., Stelmach, P., Gräßle, S., Pervan, P., et al. (2023). Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. Cell Stem Cell *30*, 706–721.e8. https://doi.org/10.1016/j.stem.2023.04.001.

17. Jakobsen, N.A., Turkalj, S., Zeng, A.G.X., Stoilova, B., Metzner, M., Nagree, M.S., Shah, S., Moore, R., Usukhbayar, B., Salazar, M.A., et al. (2023). Selective advantage of mutant stem cells in clonal hematopoiesis occurs by attenuating the deleterious effects of inflammation and aging. Preprint at bioRxiv. https://doi.org/10.1101/2023.09.12.557322.

A.5 Mutations identified by bulk targeted panels and WES at baseline (related to Chapter 3)

Patient	Mutations detected by NGS at baseline (VAF [%])	Mutation
11	NPM1 (51)	NM_002520.6(NPM1):c.863_864insCCTG p.W288fs*12
	CEBPA (48)	NM_004364.3(CEBPA):c.504dupT p.E169*
	DNMT3A (40)	NM_022552.4(DNMT3A):c.2271T>A p.N757K
	IDH1 (29)	NM_005896.2(IDH1):c.395G>A p.R132H
	NRAS (28)	NM_002524.4(NRAS):c.34G>T p.G12C
	WT1 (2)	NM_024426.4(WT1):c.1140dupG p.S381fs*4
15	ASXL1 (4)	NM_015338.5(ASXL1):c.1900_1922del p.E635fs*14
	ASXL1 (5)	NM_015338.5(ASXL1):c.1934dupG p.G646fs*11
	IDH1 (23)	NM_005896.2(IDH1):c.394C>T p.R132C
	SF3B1 (36)	NM_012433.2(SF3B1):c.2098A>G p.K700E
	TP53 (69)	NM_000546.5(TP53):c.587G>C p.R196P
5	RUNX1 (16)	NM_001754.4(RUNX1):c.485G>A p.R162K
	IDH1 (22)	NM_005896.2(IDH1):c.394C>T p.R132C
	ASXL1 (38)*	NM_015338(ASXL1):exon12:c.1888_1910del:p.H630fs
	TET2 (4)	NM_001127208.2(TET2):c.2665A>T p.K889*
20	ASXL1 (36)	NM_015338.5(ASXL1):c.1934dupG p.G646fs*11
	BRAF (2)	NM_004333.4(BRAF):c.1406G>C p.G469A
	IDH1 (42)	NM_005896.2(IDH1):c.394C>T p.R132C
	SRSF2 (45)	NM_003016.4(SRSF2):c.284C>A p.P95H
	TET2 (47)	NM_001127208.2(TET2):c.4317del p.K1439fs*9
	TET2 (46)	NM_001127208.2(TET2):c.4627A>T p.R1543*
14	DNMT3A (34)	NM_022552.4(DNMT3A):c.1627G>T p.G543C
	IDH1 (22)	NM_005896.2(IDH1):c.395G>A p.R132H
	NPM1 (38)	NM_002520.6(NPM1):c.860_863dupTCTG p.W288fs*12
9	ASXL1 (33)	NM_015338.5(ASXL1):c.1534C>T p.Q512*
	EZH2 (5)	NM_004456.4(EZH2):c.308_327dupCAGTTGCTTCAGTACCCATA p.M110fs*6
	FLT3-TKD (5)	NM_004119.2(FLT3):c.2503G>C p.D835H
	IDH1 (31)	NM_005896.2(IDH1):c.394C>T p.R132C
	STAG2 (28)	NM_006603.4(STAG2):c.3097C>T p.R1033*
	TP53 (3)	NM_000546.5(TP53):c.715A>G p.N239D
4	SRSF2 (42)	NM_003016.4(SRSF2):c.284C>A p.P95H
	ASXL2 (42)	NM_018263.4(ASXL2):c.1780C>T p.Q594*
	JAK2 (3)	NM_004972.3(JAK2):c.1849G>T p.V617F
	CBL (6)*	NM_005188(CBL):exon6:c.C895T:p.R299C
	IDH1 (38)	NM_005896.2(IDH1):c.395G>A p.R132H
	SH2B3 (34)	NM_005475.2(SH2B3):c.1226G>C p.G409A
18	ASXL1 (3)	NM_015338.5(ASXL1):c.1749G>A p.W583*
	IDH1 (17)	NM_005896.2(IDH1):c.395G>A p.R132H
	RUNX1 (29)	NM_001754.4(RUNX1):c.611G>A p.R204Q
	SRSF2 (24)	NM_003016.4(SRSF2):c.284_307del p.P95_R102del
	STAT3 (3)	NM_139276.2(STAT3):c.1940A>T p.N647I
10	BRINP3 (20)	NM_199051.1(BRINP3):c.202C>T p.R68W
	PTPN11 (7)	NM_002834.3(PTPN11):c.1508G>A p.G503E
	IDH1 (18)	NM_005896.2(IDH1):c.394C>A p.R132S
	DNMT3A (22)	NM_022552.4(DNMT3A):c.2312G>A p.R771Q
	RAD21 (27)	NM_006265.2(RAD21):c.1577_1581del p.E526fs*9
	IDH1 (45)*	IDH1:NM_001282386:exon4:c.G356C:p.R119P
	RET (58)*	RET:NM_020630:exon13:c.A2372T:p.Y791F,
	WT1 (9)*	WT1:NM_000378:exon8:c.G1334A:p.R445Q
	KRAS (47)*	KRAS:NM_033360:exon5:c.T540A:p.C180X
	KMT2D (44)*	KMT2D:NM_003482:exon39:c.T12028C:p.S4010P
	KMT2D (52)*	KMT2D:NM_003482:exon3:c.G248A:p.R83Q
	DNAH9 (50)*	DNAH9:NM_004662:exon12:c.G1576T:p.G526C
	NPM1 (27)	NM_002520.6(NPM1):c.860_863dupTCTG p.W288fs*12
	SRSF2 (28)	NM_003016.4(SRSF2):c.284C>A p.P95H

Table A.1: Mutations identified by bulk targeted panels and WES at baseline (related to Chapter 3). The asterisk in the second column denotes those mutations detected exclusively by WES. In those cases, the reported VAF was derived from WES variant calling. Otherwise, the VAF derives from targeted panels.

A.6 Mutations identified by bulk targeted panels and WES at relapse (related to Chapter 3)

Patient	[%])	Mutation
11	NPM1 (57)	NM_002520.6(NPM1):c.863_864insCCTG p.W288fs*12
	IDH1 (39)	NM_005896.2(IDH1):c.395G>A p.R132H
	DNMT3A (39)	NM_022552.4(DNMT3A):c.2271T>A p.N757K
	NRAS (45)	NM_002524.4(NRAS):c.34G>T p.G12C
	FLT3-ITD (48)*	FLT3:NM_004119:exon14:c.1817_1818insGTATGAATATGATCTCAAATGGGAGTTTCC
	CEBPA (44)	NM_004364.3(CEBPA):c.504dupT p.E169*
15	TP53 (70)	NM_000546.5(TP53):c.587G>C p.R196P
	NF1 (42)	NM_001042492.3(NF1):c.6669C>A p.C2223*
	NF1 (17)	NM_001042492.2(NF1):c.2033dupC p.I679fs*21
	SF3B1 (37)	NM_012433.3(SF3B1):c.2098A>G p.K700E
	BRINP3 (5)	NM_199051.3(BRINP3):c.107C>T p.S36L
	ASXL1 (4)	NM_015338.6(ASXL1):c.1934dupG p.G646fs*11
	ASXL1 (2)	NM_015338.6(ASXL1):c.1900_1922del p.E635fs*14
5	TET2 (1)	NM_001127208.2(TET2):c.4138C>T p.H1380Y
	TET2 (13)	NM_001127208.2(TET2):c.2665A>T p.K889*
	TET2 (2)	NM_001127208.2(TET2):c.2290del p.Q764fs*48
	SF3B1 (2)	NM_012433.3(SF3B1):c.2098A>G p.K700E
20	RUNX1 (69)	NM_001754.4(RUNX1):c.1240dupT p.Y414fs*186
	SRSF2 (48)	NM_003016.4(SRSF2):c.284C>A p.P95H
	TET2 (46)	NM_001127208.2(TET2):c.4317del p.K1439fs*9
	TET2 (45)	NM_001127208.2(TET2):c.4627A>T p.R1543*
	ASXL1 (8)	NM_015338.6(ASXL1):c.1934dupG p.G646fs*11
14	CEBPA (1)	NM_004364.3(CEBPA):c.68dupC p.H24fs*84
	NPM1 (49)	NM_002520.6(NPM1):c.860_863dupTCTG p.W288fs*12
	FBXW7 (1)	NM_033632.3(FBXW7):c.2065C>T p.R689W
	IDH1 (1)*	IDH1:NM_001282386:exon4:c.G395A:p.R132H
	DNMT3A (39)	NM_022552.4(DNMT3A):c.1627G>T p.G543C
9	ASXL1 (41)	NM_015338.6(ASXL1):c.1534C>T p.Q512*
	IDH1 (24)	NM_005896.3(IDH1):c.394C>T p.R132C
	CEBPA (22)	NM_004364.4(CEBPA):c.765del p.L256fs*62
	CEBPA (13)	NM_004364.4(CEBPA):c.174_184del p.E59fs*45
	EZH2 (12)	NM_004456.5(EZH2):c.308_327dupCAGTTGCTTCAGTACCCATA p.M110fs*6
	NRAS (12)	NM_002524.5(NRAS):c.37G>C p.G13R
	RUNX1 (12)	NM_001754.4(RUNX1):c.911dupC p.G305fs*295
	STAG2 (4)	NM_006603.5(STAG2):c.1002G>A p.W334*
	STAG2 (4)	NM_006603.5(STAG2):c.3097C>T p.R1033*
	RUNX1 (3)	NM_001754.4(RUNX1):c.602G>A p.R201Q
	FLT3-TKD (3)	NM_004119.3(FLT3):c.2503G>C p.D835H
	PTPN11 (2)	NM_002834.4(PTPN11):c.178G>C p.G60R
	TP53 (2)	NM_000546.5(TP53):c.715A>G p.N239D
4	ETV6 (2)	NM_001987.5(ETV6):c.775dupC p.R259fs*41
	TET2 (1)	NM_001127208.2(TET2):c.5698G>C p.V1900L
	ETV6 (3)	NM_001987.5(ETV6):c.1075C>G p.R359G
	JAK2 (2)	NM_004972.3(JAK2):c.1849G>T p.V617F
	SH2B3 (28)	NM_005475.3(SH2B3):c.1226G>C p.G409A
	IDH1 (29)	NM_005896.3(IDH1):c.395G>A p.R132H
	ASXL2 (32)	NM_018263.6(ASXL2):c.1780C>T p.Q594*
	SRSF2 (33)	NM_003016.4(SRSF2):c.284C>A p.P95H
10	GNAS (1) (from a remission time point)	NM_000516.6(GNAS):c.602G>A p.R201H

Table A.2: Mutations identified by bulk targeted panels and WES at relapse (related to Chapter 3). The asterisk in the second column denotes those mutations detected exclusively by WES. In those cases, the reported VAF was derived from WES variant calling. Otherwise, the VAF derives from targeted panels. Note that the results for pt10 (who remains in sustained remission) were obtained from a longer-term remission rather than relapse time point.

A.7 Pre-amplification primers used for TARGET-seq+ (related to Chapter 3)

Mutation	Amplicon	Forward pre-amplification primer sequence	Reverse pre-amplification primer sequence	Patients
SRSF2-P95H	gDNA	CAAGGTGGACAACCTGACCTAC	CAAGCACAGCGGGGTTAATTC	pt4
ASXL2-Q594*	gDNA	TGCTTCTGCATTAAGAACTACCAC	TCAGGGGTTCTCCTAGGCATT	pt4
IDH1-R132H/*IDH1*-R132C	gDNA	TCATACCTTGCTTAATGGGTGT	TGGTGTACTCAGAGCCTTCG	pt4, pt9, pt11, pt14, pt15
JAK2-V617F	gDNA	ccaaagcacattgtatcctcatct	cactgacacctagctgtgatcct	pt4
CBL-R299C	gDNA	CTTGCCTTCCACCGTAATACCA	CCAGGACCTGAAGGTCAAATCAA	pt4
ASXL1-Q512*	gDNA	TACAAGGATGGGGAGGCTAAGA	TCTCATGACCCTCAAAGAAAACCT	pt9
STAG2-R1033*	gDNA	GCCTATGCTCGCACAACTAT	AACCAACTCTGGAGTGCTAGG	pt9
FLT3-D835H	gDNA	CTCTGAGATTCCCTCTGAAGGC	ACGACACAACACAAAATAGCCG	pt9
NRAS-G13R	gDNA	AATCCGGTGTTTTTGCGTTCTC	CAGTCTCGCTACTATGGCCTG	pt9, pt11
NPM1-W288fs*12	gDNA	GTAACATTGGCCATATGGGTCTC	TTTGGACAACACATTCTTGGCA	pt11, pt14
DNMT3A-N757K/*DNMT3A*-R771Q	gDNA	GGTTCTTTCCGAGGTAGGCC	CTGGGTCTCCTCTCTTTCGTG	pt11, pt10
IDH1-R132H	gDNA	TGTTGAGATGGACGCCTATTTGT	TGGTGTACTCAGAGCCTTCG	pt11, pt14
FLT3-ITD	gDNA	AGAACTGCCTATTCCTAACTGACT	CAAATGGTGAGTACGTGCATTTT	pt11
DNMT3A-G543C	gDNA	CCAGCTAAGGAGACCACTGG	GGGTCATGTCTTCAGGGCTT	pt14
TP53-R196P	gDNA	TCTGATTCCTCACTGATTGCTCTT	TTTACTTTGCACATCTCATGGGG	pt15
NF1-C2223*	gDNA	TTGCATGGACTGTGTTATTGGT	TGCAATTAAAAGATCCACAGAACT	pt15
NF1-I679fs*21	gDNA	ACGTTATGGTGTTTACCTGTCCA	AATACATTCTCAAAATCAGATGCCA	pt15
SF3B1-K700E	gDNA	CTGCAGTTTGGCTGAATAGTTGA	ATCAACTGACCTGAAATGAAGAGAA	pt5, pt15
IDH1-R132H/*IDH1*-R132C/*IDH1*-R132S	gDNA	TGTTGAGATGGACGCCTATTTGT	GTTTAGGGTGTGCCAGTGCT	pt5, pt10, pt18, pt20
RUNX1-R162K	gDNA	TGCATTTGTCCTTTGACTGGTG	AGACAGACCGAGTTTCTAGGG	pt5
TET2-K889*	gDNA/cDNA	ACCAGGTGCACTTCTCCAAAA	GCAGCTTGTTGACCAGACAT	pt5
ASXL1-H630fs/*ASXL1*-G646fs*11/*ASXL1*-W583*	gDNA	TGCTTTACAGTCCCTAGGTCAG	TCGCTGTAGATCTGACGTACAC	pt5, pt18, pt20
BRINP3-R68W	gDNA	TTAGCCACCTCAACACCTTAACTG	GGTGCATGGTAAATAACAACCC	pt10
PTPN11-G503E	gDNA	TTAGCATTGTCTCTGAGTCCAC	TCCTGCGCTGTAGTGTTTCA	pt10
RAD21-E526fs*9	gDNA	GGACCCAGAAAAGCAATGTCC	ATAAAGCTGGTTGGAGGTTTTTGC	pt10
NPM1-W288fs*12	gDNA	TGTCTATGAAGTGTTGTGGTTCC	ACCACCGCTACTACTATGTCC	pt10
SRSF2-P95H/*SRSF2*-P95_R102del	gDNA	GACCGCTACACCAAGGAGTC	AGACGCCATTTCCCCAGTC	pt10, pt18, pt20
WT1-R462Q	gDNA	TGGGGGACTGGGGAAATCTAA	TCCTTCTCTCAACTGAGTCTAAACC	pt10
KRAS-C180X	gDNA	TTACTCTTACCAATGCAACAGACT	GCTTTCTGCCAAAATTAATGTGCT	pt10
RUNX1-R204Q	gDNA	AGCCAAAATTCCGGGAGTGTT	CATGGTCCCTGAGTAtaccagc	pt18
BRAF-G469A	gDNA	AGGGGATCTCTTCCTGTATCC	AGGACTAGTTAACCTGGAGGAATTC	pt20
TET2-K1439fs*9	gDNA	TGGGTGGGCAATGTCAATTTG	GTTTAGCCTGGCTTGCGTTT	pt20
TET2-R1543*	gDNA	TCTTACCCTGTCCACAGAACTT	ATATGGGGAGCAGTTGTCCAC	pt20
RUNX1-Y414fs*186	gDNA	CTCCGCAACCTCCTACTCAC	GGAATCTTCCTGTTTGCTTTCC	pt20
SRSF2-P95H	cDNA	CTTCGTGCGGATCTGGACTT	CTTCGTTCGCTTTCACGACAA	pt4
ASXL2-Q594*	cDNA	CAGCCGAACTTCTCCACATCT	ACCAGTTGGGCTTTTGCTTT	pt4
IDH1-R132H	cDNA	AATCAGTGGCGGTTCTGTGG	CCGTCACTTGGTGTGTAGGT	pt4, pt9, pt11, pt14, pt15
JAK2-V617F	cDNA	TAAATGCTGTCCCCCAAAGC	CCATGCCAACTGTTTAGCAAC	pt4
CBL-R299C	cDNA	GCTCGGCTCCAGAAATTCATTC	TGATATGGTCTTGGGGAGTTGG	pt4
ASXL1-Q512*	cDNA	TGCAAGTCAGGATACGACAGG	TCCGGGGGCATATCTGGTAA	pt9
STAG2-R1033*	cDNA	AAAACAAGAGAAGCCATTGCCA	GCTCATTGCTGCTCTCTCCTG	pt9
FLT3-D835H	cDNA	GCTGGAAGAAGAGGAGGACTTG	GTGTAGATGCCTTCAAACAGGC	pt9
NRAS-G13R	cDNA	GCGCCGACTGATTACGTAGC	GGCAAATACACAGAGGAAGCCT	pt9, pt11
NPM1-W288fs*12	cDNA	GTGGTTCTCTTCCCAAAGTGGA	CCATGTCTGACCACCGCTA	pt11, pt14
DNMT3A-N757K/*DNMT3A*-R771Q	cDNA	GTGGACCGCTACATTGCCTC	TGGTCCTCACTTTGCTGAACT	pt11, pt10
DNMT3A-G543C	cDNA	TGGGAGCCTCAATGTTACCC	TGTAGCGGTCCACCTGAATG	pt14
TP53-R196P	cDNA	ATCTACAAGCAGTCACAGCACA	AGTGTGATGATGGTGAGGATGG	pt15
IDH1-R132H/*IDH1*-R132S	cDNA	AATCAGTGGCGGTTCTGTGG	GGATTGGTGGACGTCTCCTG	pt5, pt10, pt18, pt20
RUNX1-R162K	cDNA	AACTTCCTCTGCTCCGTGC	CTTGCGGTGGGTTTGTGAAG	pt5
ASXL1-H630fs	cDNA	AGCCACAGCCCACTAAAGAG	TCGCTGTAGATCTGACGTACAC	pt5, pt18, pt20
SF3B1-K700E	cDNA	TGTGCCATCTTGCCACATCT	CTGTGTTGGCGGATACCCTT	pt5, pt15
PTPN11-G503E	cDNA	CCGGACAGGGACGTTCATTG	TCCTGCGCTGTAGTGTTTCA	pt10
NPM1-W288fs*12	cDNA	CTTCCCAAAGTGGAAGCCAA	ACCACCGCTACTACTATGTCC	pt10
RUNX1-R204Q	cDNA	TCACTGTGATGGCTGGCAAT	CTGATCGTAGGACCACGGTG	pt18
RUNX1-Y414fs*186	cDNA	CCTCAGCCTCAGAGTCAGATG	GGAATCTTCCTGTTTGCTTTCC	pt20

Table A.3: Pre-amplification primers used for TARGET-seq+ (related to Chapter 3). When a primer pair was used to screen more than one patient, all patients for which that pair was used are listed. Note that all mutations were genotyped from gDNA, but only some were genotyped from cDNA.

A.8 Nested (PCR1) primers used for TARGET-seq+ (related to Chapter 3)

Mutation	Amplicon	Forward pre-amplification primer sequence	Reverse pre-amplification primer sequence	Patients
SRSF2-P95H	gDNA	CAAGGTGGACAACCTGACCTAC	CAAGCACAGCGGGGTTAATTC	pt4
ASXL2-Q594*	gDNA	TGCTTCTGCATTAAGAACTACCAC	TCAGGGGTTCTCCTAGGCATT	pt4
IDH1-R132H/*IDH1*-R132C	gDNA	TCATACCTTGCTTAATGGGTGT	TGGTGTACTCAGAGCCTTCG	pt4, pt9, pt11, pt14, pt15
JAK2-V617F	gDNA	ccaaagcacattgtatcctcatct	cactgacacctagctgtgatcct	pt4
CBL-R299C	gDNA	CTTGCCTTCCACCGTAATACCA	CCAGGACCTGAAGGTCAAATCAA	pt4
ASXL1-Q512*	gDNA	TACAAGGATGGGGAGGCTAAGA	TCTCATGACCCTCAAAGAAAACCT	pt9
STAG2-R1033*	gDNA	GCCTATGCTCGCACAACTAT	AACCAACTCTGGAGTGCTAGG	pt9
FLT3-D835H	gDNA	CTCTGAGATTCCCTCTGAAGGC	ACGACACAACACAAAATAGCCG	pt9
NRAS-G13R	gDNA	AATCCGGTGTTTTTGCGTTCTC	CAGTCTCGCTACTATGGCCTG	pt9, pt11
NPM1-W288fs*12	gDNA	GTAACATTGGCCATATGGGTCTC	TTTGGACAACACATTCTTGGCA	pt11, pt14
DNMT3A-N757K/*DNMT3A*-R771Q	gDNA	GGTTCTTTCCGAGGTAGGCC	CTGGGTCTCCTCTCTTTCGTG	pt11, pt10
IDH1-R132H	gDNA	TGTTGAGATGGACGCCTATTTGT	TGGTGTACTCAGAGCCTTCG	pt11, pt14
FLT3-ITD	gDNA	AGAACTGCCTATTCCTAACTGACT	CAAATGGTGAGTACGTGCATTTT	pt11
DNMT3A-G543C	gDNA	CCAGCTAAGGAGACCACTGG	GGGTCATGTCTTCAGGGCTT	pt14
TP53-R196P	gDNA	TCTGATTCCTCACTGATTGCTCTT	TTTACTTTGCACATCTCATGGGG	pt15
NF1-C2223*	gDNA	TTGCATGGACTGTGTTATTGGT	TGCAATTAAAAGATCCACAGAACT	pt15
NF1-I679fs*21	gDNA	ACGTTATGGTGTTTACCTGTCCA	AATACATTCTCAAAATCAGATGCCA	pt15
SF3B1-K700E	gDNA	CTGCAGTTTGGCTGAATAGTTGA	ATCAACTGACCTGAAATGAAGAGAA	pt5, pt15
IDH1-R132H/*IDH1*-R132C/*IDH1*-R132S	gDNA	TGTTGAGATGGACGCCTATTTGT	GTTTAGGGTGTGCCAGTGCT	pt5, pt10, pt18, pt20
RUNX1-R162K	gDNA	TGCATTTGTCCTTTGACTGGTG	AGACAGACCGAGTTTCTAGGG	pt5
TET2-K889*	gDNA/cDNA	ACCAGGTGCACTTCTCCAAAA	GCAGCTTGTTGACCAGACAT	pt5
ASXL1-H630fs/*ASXL1*-G646fs*11/*ASXL1*-W583*	gDNA	TGCTTTACAGTCCCTAGGTCAG	TCGCTGTAGATCTGACGTACAC	pt5, pt18, pt20
BRINP3-R68W	gDNA	TTAGCCACCTCAACACCTTAACTG	GGTGCATGGTAAATAACAACCC	pt10
PTPN11-G503E	gDNA	TTAGCATTGTCTCTGAGTCCAC	TCCTGCGCTGTAGTGTTTCA	pt10
RAD21-E526fs*9	gDNA	GGACCCAGAAAAGCAATGTCC	ATAAAGCTGGTTGGAGGTTTTTGC	pt10
NPM1-W288fs*12	gDNA	TGTCTATGAAGTGTTGTGGTTCC	ACCACCGCTACTACTATGTCC	pt10
SRSF2-P95H/*SRSF2*-P95_R102del	gDNA	GACCGCTACACCAAGGAGTC	AGACGCCATTTCCCCAGTC	pt10, pt18, pt20
WT1-R462Q	gDNA	TGGGGGACTGGGGAAATCTAA	TCCTTCTCTCAACTGAGTCTAAACC	pt10
KRAS-C180X	gDNA	TTACTCTTACCAATGCAACAGACT	GCTTTCTGCCAAAATTAATGTGCT	pt10
RUNX1-R204Q	gDNA	AGCCAAAATTCCGGGAGTGTT	CATGGTCCCTGAGTAtaccagc	pt18
BRAF-G469A	gDNA	AGGGGATCTCTTCCTGTATCC	AGGACTAGTTAACCTGGAGGAATTC	pt20
TET2-K1439fs*9	gDNA	TGGGTGGGCAATGTCAATTTG	GTTTAGCCTGGCTTGCGTTT	pt20
TET2-R1543*	gDNA	TCTTACCCTGTCCACAGAACTT	ATATGGGGAGCAGTTGTCCAC	pt20
RUNX1-Y414fs*186	gDNA	CTCCGCAACCTCCTACTCAC	GGAATCTTCCTGTTTGCTTTCC	pt20
SRSF2-P95H	cDNA	CTTCGTGCGGATCTGGACTT	CTTCGTTCGCTTTCACGACAA	pt4
ASXL2-Q594*	cDNA	CAGCCGAACTTCTCCACATCT	ACCAGTTGGGCTTTTGCTTT	pt4
IDH1-R132H	cDNA	AATCAGTGGCGGTTCTGTGG	CCGTCACTTGGTGTGTAGGT	pt4, pt9, pt11, pt14, pt15
JAK2-V617F	cDNA	TAAATGCTGTCCCCCAAAGC	CCATGCCAACTGTTTAGCAAC	pt4
CBL-R299C	cDNA	GCTCGGCTCCAGAAATTCATTC	TGATATGGTCTTGGGGAGTTGG	pt4
ASXL1-Q512*	cDNA	TGCAAGTCAGGATACGACAGG	TCCGGGGGCATATCTGGTAA	pt9
STAG2-R1033*	cDNA	AAAACAAGAGAAGCCATTGCCA	GCTCATTGCTGCTCTCTCCTG	pt9
FLT3-D835H	cDNA	GCTGGAAGAAGAGGAGGACTTG	GTGTAGATGCCTTCAAACAGGC	pt9
NRAS-G13R	cDNA	GCGCCGACTGATTACGTAGC	GGCAAATACACAGAGGAAGCCT	pt9, pt11
NPM1-W288fs*12	cDNA	GTGGTTCTCTTCCCAAAGTGGA	CCATGTCTGACCACCGCTA	pt11, pt14
DNMT3A-N757K/*DNMT3A*-R771Q	cDNA	GTGGACCGCTACATTGCCTC	TGGTCCTCACTTTGCTGAACT	pt11, pt10
DNMT3A-G543C	cDNA	TGGGAGCCTCAATGTTACCC	TGTAGCGGTCCACCTGAATG	pt14
TP53-R196P	cDNA	ATCTACAAGCAGTCACAGCACA	AGTGTGATGATGGTGAGGATGG	pt15
IDH1-R132H/*IDH1*-R132S	cDNA	AATCAGTGGCGGTTCTGTGG	GGATTGGTGGACGTCTCCTG	pt5, pt10, pt18, pt20
RUNX1-R162K	cDNA	AACTTCCTCTGCTCCGTGC	CTTGCGGTGGGTTTGTGAAG	pt5
ASXL1-H630fs	cDNA	AGCCACAGCCCACTAAAGAG	TCGCTGTAGATCTGACGTACAC	pt5, pt18, pt20
SF3B1-K700E	cDNA	TGTGCCATCTTGCCACATCT	CTGTGTTGGCGGATACCCTT	pt5, pt15
PTPN11-G503E	cDNA	CCGGACAGGGACGTTCATTG	TCCTGCGCTGTAGTGTTTCA	pt10
NPM1-W288fs*12	cDNA	CTTCCCAAAGTGGAAGCCAA	ACCACCGCTACTACTATGTCC	pt10
RUNX1-R204Q	cDNA	TCACTGTGATGGCTGGCAAT	CTGATCGTAGGACCACGGTG	pt18
RUNX1-Y414fs*186	cDNA	CCTCAGCCTCAGAGTCAGATG	GGAATCTTCCTGTTTGCTTTCC	pt20

Table A.4: Nested (PCR1) primers used for TARGET-seq+ (related to Chapter 3). When a primer pair was used to screen more than one patient, all patients for which that pair was used are listed. Note that all mutations were genotyped from gDNA, but only some were genotyped from cDNA.

A.9 Rates of successful clonal assignment per sample (related to Chapter 3)

Sample	Single cells sorted	Single cells passing genotyping QC	Success rate (%)
pt4-BL	1056	939	89
pt4-EOC3	528	477	90
pt4-REL	1056	924	88
pt9-BL	1232	974	79
pt9-EOC3	680	633	93
pt9-REL	1408	1278	91
pt11-BL	880	746	85
pt11-EOC3	704	526	75
pt11-REL	1232	1199	97
pt14-BL	704	286	41
pt14-EOC3	704	427	61
pt14-REL	880	583	66
pt15-BL	1056	936	89
pt15-EOC3	528	52	10
pt15-REL	477	27	6
pt5-BL	1408	1229	87
pt5-EOC3	704	176	25
pt5-EOC11	688	178	26
pt10-BL	609	374	61
pt10-EOC3	528	431	82
pt10-EOC11	745	641	86
pt18-BL	856	679	79
pt18-EOC1	609	364	60
pt18-EOC9	669	180	27
pt20-BL	704	439	62
pt20-EOC1	704	296	42
pt20-EOC11	662	404	61

Table A.5: Rates of successful clonal assignment per sample (related to Chapter 3). Success rate represents the ratio of cells assigned to genetic clones versus total number of cells sorted per sample.

A.10 inferCNV genome-wide for pt15 (related to Chapter 3)

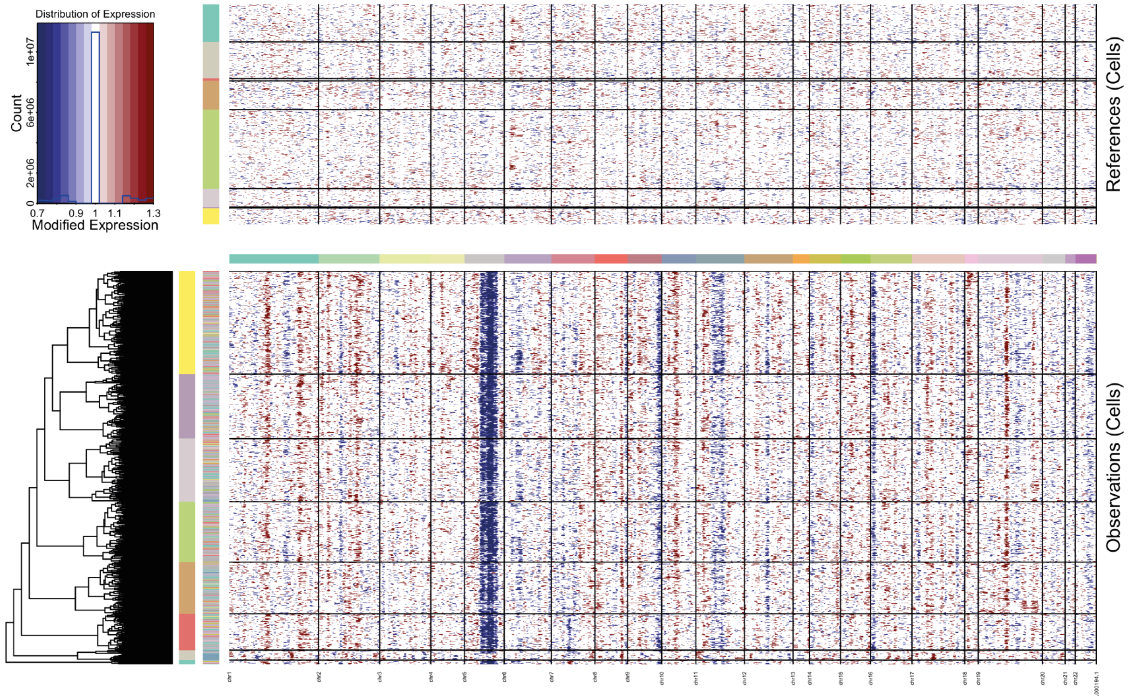


Figure A.1: inferCNV genome-wide for pt15 (related to Chapter 3). The upper heatmap shows inferCNV results from WT control BM cells for which no CNAs were expected. The bottom heatmap shows inferCNV results from all cells from pt15 which passed scRNA-seq QC.

A.11 inferCNV genome-wide for pt5 (related to Chapter 3)

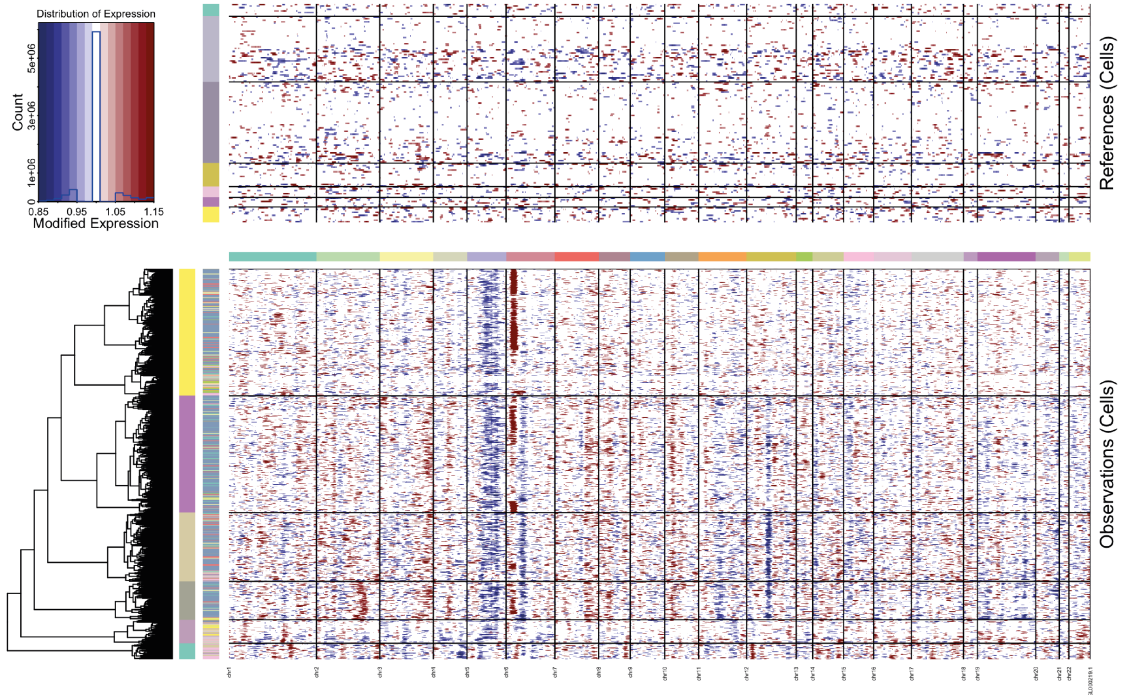


Figure A.2: inferCNV genome-wide for pt5 (related to Chapter 3). The upper heatmap shows inferCNV results from WT control BM cells for which no CNAs were expected. The bottom heatmap shows inferCNV results from all cells from pt5 which passed scRNA-seq QC.

A.12 inferCNV genome-wide for pt14 (related to Chapter 3)

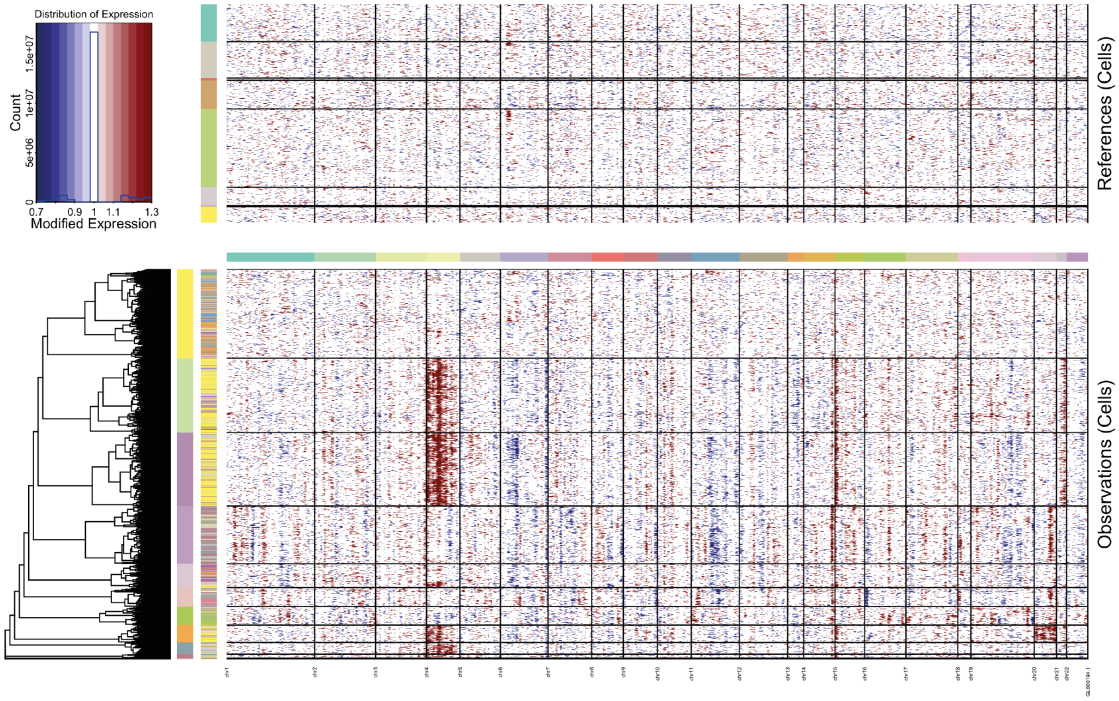


Figure A.3: inferCNV genome-wide for pt14 (related to Chapter 3). The upper heatmap shows inferCNV results from WT control BM cells for which no CNAs were expected. The bottom heatmap shows inferCNV results from all cells from pt14 which passed scRNA-seq QC.

A.13 inferCNV genome-wide for pt9 (related to Chapter 3)

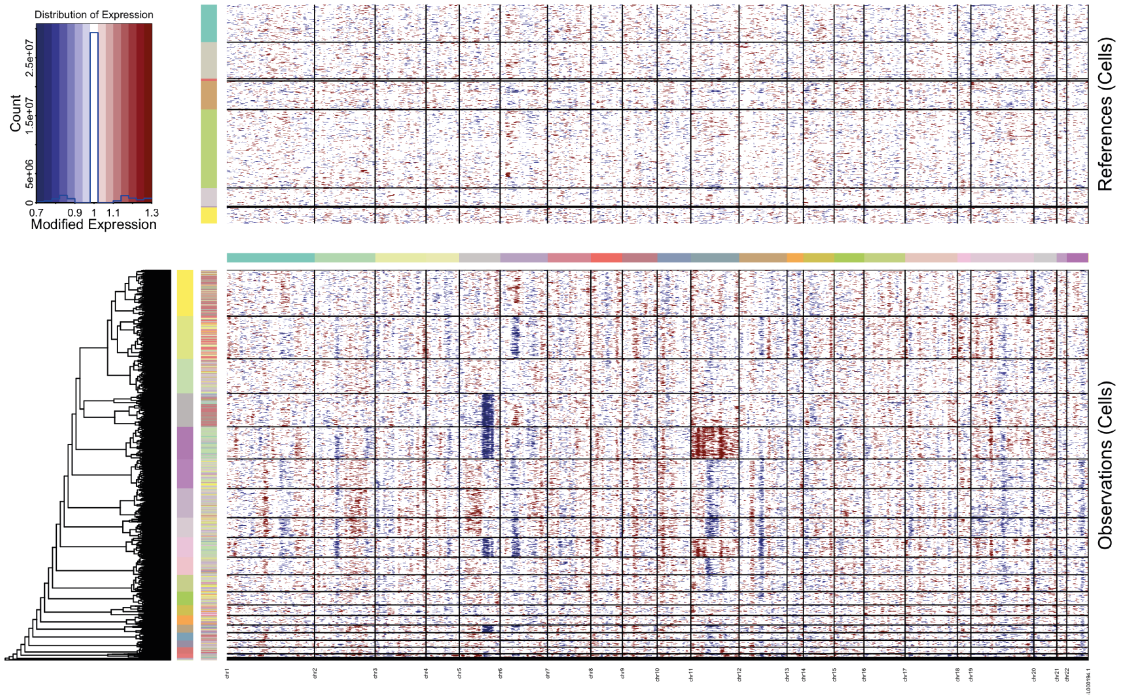


Figure A.4: inferCNV genome-wide for pt9 (related to Chapter 3). The upper heatmap shows inferCNV results from WT control BM cells for which no CNAs were expected. The bottom heatmap shows inferCNV results from all cells from pt9 which passed scRNA-seq QC.

A.14 inferCNV genome-wide for pt4 (related to Chapter 3)

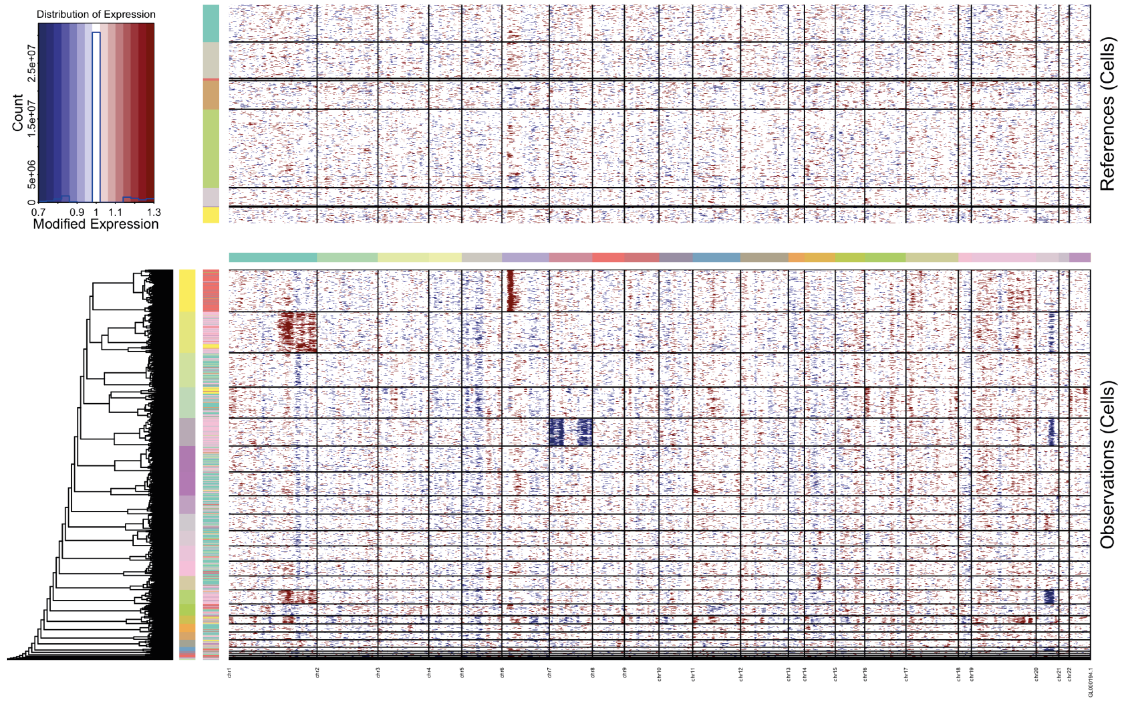


Figure A.5: inferCNV genome-wide for pt4 (related to Chapter 3). The upper heatmap shows inferCNV results from WT control BM cells for which no CNAs were expected. The bottom heatmap shows inferCNV results from all cells from pt4 which passed scRNA-seq QC.

A.15 Inferred clone sizes within all live, Lin$^-$ cells for pt11 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt11-BL	DNINF	0	0	0.046161326
pt11-BL	DNIN	0.407472529	0.315856916	0.509731357
pt11-BL	DNI	0.32897037	0.244730286	0.43107534
pt11-BL	DN	0.063518833	0.037716743	0.130941915
pt11-BL	WT	0.053362139	0.022374808	0.12101876
pt11-BL	N/DN	0.131368037	0.076759107	0.219065065
pt11-BL	D/DN	0.012683487	0.002381405	0.067744392
pt11-EOC3	DNINF	0.240772029	0.174875368	0.330804028
pt11-EOC3	DNIN	0.024792492	0.007823845	0.08106028
pt11-EOC3	DNI	0.022217623	0.00615333	0.077662782
pt11-EOC3	DN	0	3.12E-18	0.041242825
pt11-EOC3	WT	0.623768196	0.526113353	0.708418312
pt11-EOC3	N/DN	0.064428261	0.029818956	0.134785077
pt11-EOC3	D/DN	0.021395349	0.005888852	0.076315018
pt11-REL	DNINF	0.853005776	0.779594373	0.90577928
pt11-REL	DNIN	0.058602345	0.030424201	0.116289532
pt11-REL	DNI	0.025801975	0.009463171	0.073835609
pt11-REL	DN	0.001035027	1.91E-04	0.035630012
pt11-REL	WT	0.024677419	0.00846176	0.071781356
pt11-REL	N/DN	0.036795396	0.014446424	0.089010668
pt11-REL	D/DN	0	3.96E-19	0.03376784

Table A.6: Inferred clone sizes within all live, Lin$^-$ cells for pt11 (related to Chapter 3). The inferred size of each clone at each time point is shown in the third column. For details about the analysis, please refer to Materials and Methods (section 2.8.7)

A.16 Inferred clone sizes within all live, Lin$^-$ cells for pt5 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt5-BL	WT	0.125436817	0.07822475	0.20143922
pt5-BL	del5	0.001043196	2.86E-04	0.051628916
pt5-BL	5S	0	0	0.049838854
pt5-BL	T	0.055609756	0.029168147	0.122362314
pt5-BL	5RA	0.162120691	0.107641544	0.250903606
pt5-BL	5RAI	0.605022896	0.519317182	0.691287885
pt5-BL	5R	0.050766644	0.0280028	0.119092855
pt5-EOC3	WT	0.911201224	0.784711243	0.965135968
pt5-EOC3	del5	0.00938604	0.004439643	0.101549338
pt5-EOC3	5S	9.54E-04	3.53E-04	0.091354462
pt5-EOC3	T	0.077834235	0.027124108	0.204186866
pt5-EOC3	5RA	0	6.10E-18	0.090052332
pt5-EOC3	5RAI	0	6.10E-18	0.090052332
pt5-EOC3	5R	0	6.10E-18	0.090052332
pt5-EOC11	WT	0.632466295	0.474022763	0.754803589
pt5-EOC11	del5	0.106999374	0.04624515	0.245434602
pt5-EOC11	5S	0.071165707	0.02684828	0.198488565
pt5-EOC11	T	0.184438182	0.099936559	0.335378826
pt5-EOC11	5RA	4.94E-04	8.78E-05	0.096198455
pt5-EOC11	5RAI	0	1.04E-19	0.095438845
pt5-EOC11	5R	0	1.04E-19	0.095438845

Table A.7: Inferred clone sizes within all live, Lin$^-$ cells for pt5 (related to Chapter 3). As in Table A.6 but for pt5.

A.17 Inferred clone sizes within all live, Lin$^-$ cells for pt14 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt14-BL	DNI	0.547488323	0.430827857	0.662486026
pt14-BL	DN	0.150179987	0.084530045	0.256929197
pt14-BL	DN-4	0.043682516	0.014983324	0.125179182
pt14-BL	DN-4/21	0	5.87E-20	0.057502501
pt14-BL	DN-4/20	0	5.87E-20	0.057502501
pt14-BL	D	0.045823521	0.015938153	0.127942705
pt14-BL	WT	0.212825653	0.131216397	0.322474619
pt14-EOC3	DNI	0.017613636	0.003116205	0.094629342
pt14-EOC3	DN	0.002124058	9.63E-04	0.067907529
pt14-EOC3	DN-4	0.057761662	0.029779982	0.144068736
pt14-EOC3	DN-4/21	0.035352547	0.009745662	0.121136856
pt14-EOC3	DN-4/20	0.001145545	3.72E-04	0.066615466
pt14-EOC3	D	0.159675355	0.086468796	0.276864759
pt14-EOC3	WT	0.726327196	0.600731765	0.821611242
pt14-REL	DNI	0	0	0.042139378
pt14-REL	DN	0.055326196	0.02549408	0.117380611
pt14-REL	DN-4	0.54432248	0.476283061	0.626503894
pt14-REL	DN-4/21	0.190738362	0.157913495	0.260396681
pt14-REL	DN-4/20	0.091790387	0.050168056	0.159513411
pt14-REL	D	0.024082179	0.008600075	0.078316968
pt14-REL	WT	0.091911635	0.048666096	0.155663558

Table A.8: Inferred clone sizes within all live, Lin$^-$ cells for pt14 (related to Chapter 3). As in Table A.6 but for pt14.

A.18 Inferred clone sizes within all live, Lin$^-$ cells for pt9 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt9-BL	AISF	0.100735936	0.068610053	0.151999086
pt9-BL	AIS	0.470199619	0.419881587	0.520743605
pt9-BL	AI	0.127977829	0.096841404	0.177837078
pt9-BL	AI-5	0.020194357	0.011652	0.052853918
pt9-BL	AI-5/11	0.182961613	0.158882847	0.226464138
pt9-BL	AN	0	0	0.025054392
pt9-BL	A	0.076338347	0.051767281	0.122486905
pt9-BL	WT	0.021504418	0.01074953	0.054326764
pt9-EOC3	AISF	0.043024649	0.014629927	0.165160559
pt9-EOC3	AIS	0.057096879	0.015714422	0.188195653
pt9-EOC3	AI	0.003080178	8.95E-04	0.106220605
pt9-EOC3	AI-5	0.072708779	0.026396211	0.206202994
pt9-EOC3	AI-5/11	0.08696467	0.031289118	0.22671243
pt9-EOC3	AN	0.03026804	0.006023349	0.150057864
pt9-EOC3	A	0.194160206	0.098566372	0.353107885
pt9-EOC3	WT	0.512559152	0.355259869	0.664611674
pt9-REL	AISF	0.72746299	0.677487314	0.774080572
pt9-REL	AIS	0.087832592	0.060593242	0.127752536
pt9-REL	AI	0.029580551	0.015876452	0.057107546
pt9-REL	AI-5	0.044942099	0.027211346	0.075627808
pt9-REL	AI-5/11	0.004534298	0.001090132	0.023160353
pt9-REL	AN	0.014186934	0.007866111	0.035178731
pt9-REL	A	0.08098435	0.059990512	0.113397387
pt9-REL	WT	0.010138846	0.004038367	0.030981247

Table A.9: Inferred clone sizes within all live, Lin$^-$ cells for pt9 (related to Chapter 3). As in Table A.6 but for pt9.

A.19 Inferred clone sizes within all live, Lin$^-$ cells for pt4 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt4-BL	SIAC	0.183777366	0.087328369	0.343648308
pt4-BL	SIAJ	0.269646892	0.245822619	0.380317062
pt4-BL	SIAJ-1	0.004152182	0.002226954	0.122579059
pt4-BL	SIA-20/1	0.03	0.0053245	0.162881025
pt4-BL	SIA-20/7	7.75E-04	2.05E-04	0.118126097
pt4-BL	SIA	0.177771414	0.085141914	0.340317593
pt4-BL	SIA-20	0.002798666	0.001420711	0.120940196
pt4-BL	SI	0.060774732	0.016889419	0.2042679
pt4-BL	S	0.270150132	0.152928848	0.426874091
pt4-BL	WT	0	0	0.116792521
pt4-EOC3	SIAC	0.273929258	0.123047414	0.496865395
pt4-EOC3	SIAJ	0.108887857	0.06007016	0.308940575
pt4-EOC3	SIAJ-1	0.001851972	8.00E-04	0.177028375
pt4-EOC3	SIA-20/1	0.055925946	0.010710259	0.255628983
pt4-EOC3	SIA-20/7	0.053440157	0.00950284	0.25209336
pt4-EOC3	SIA	0.014560562	0.008418782	0.190699194
pt4-EOC3	SIA-20	0.002001804	6.14E-04	0.177018244
pt4-EOC3	SI	0.110167211	0.031623385	0.323386902
pt4-EOC3	S	0.272679115	0.123588943	0.496514117
pt4-EOC3	WT	0.106470588	0.029756195	0.318281065
pt4-REL	SIAC	0.106870613	0.060327872	0.18843783
pt4-REL	SIAJ	0.194340795	0.156421952	0.280394609
pt4-REL	SIAJ-1	0.091880187	0.06098432	0.170463131
pt4-REL	SIA-20/1	0.191081039	0.142032943	0.278436764
pt4-REL	SIA-20/7	0.125533632	0.100634088	0.205026462
pt4-REL	SIA	0.141470065	0.09714435	0.229839338
pt4-REL	SIA-20	0.016729705	0.006912797	0.083407699
pt4-REL	SI	0.022982338	0.011109846	0.091522615
pt4-REL	S	0.097086902	0.045425771	0.175923048
pt4-REL	WT	0.012024725	0.003296419	0.076730744

Table A.10: Inferred clone sizes within all live, Lin$^-$ cells for pt4 (related to Chapter 3). As in Table A.6 but for pt4.

A.20 Inferred clone sizes within all live, Lin$^-$ cells for pt18 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt18-BL	WT	0.17610357	0.10452245	0.272838851
pt18-BL	A	0.065467986	0.025409074	0.159785689
pt18-BL	R	0.04301545	0.013927118	0.134897032
pt18-BL	RI	0.263215084	0.172954849	0.380583377
pt18-BL	RIS	0.45219791	0.387046213	0.55988753
pt18-EOC1	WT	0.47703278	0.3942371	0.562051457
pt18-EOC1	A	0.258866562	0.192379611	0.341332111
pt18-EOC1	R	0.00707483	0.001250097	0.043615409
pt18-EOC1	RI	0.086735081	0.050086044	0.149208072
pt18-EOC1	RIS	0.170043834	0.115931276	0.242801441
pt18-EOC9	WT	0.719121015	0.617564115	0.80224597
pt18-EOC9	A	0.258268708	0.178323414	0.358739128
pt18-EOC9	R	0	7.34E-20	0.044673687
pt18-EOC9	RI	0.021020408	0.005644474	0.078976848
pt18-EOC9	RIS	0	7.34E-20	0.044673687

Table A.11: Inferred clone sizes within all live, Lin$^-$ cells for pt18 (related to Chapter 3). As in Table A.6 but for pt18.

A.21 Inferred clone sizes within all live, Lin$^-$ cells for pt10 (related to Chapter 3)

Sample	Clone	Estimated frequency	Lower limit of the CI	Upper limit of the CI
pt10-BL	BSNDKIRI	0.269670675	0.239614794	0.32678524
pt10-BL	WT	0.6971753	0.649646228	0.725189317
pt10-BL	BSNDKIRIP	0.015522388	0.008964314	0.06214752
pt10-BL	BSNDKIW	0.007105089	0.002848116	0.045049877
pt10-EOC3	BSNDKIRI	0	0	0.022848445
pt10-EOC3	WT	0.999219114	0.976370669	0.999219114
pt10-EOC3	BSNDKIRIP	0	0	0.022848445
pt10-EOC3	BSNDKIW	0	0	0.022848445
pt10-EOC11	BSNDKIRI	0	3.03E-20	0.029760781
pt10-EOC11	WT	0.999500943	0.969740161	0.999500943
pt10-EOC11	BSNDKIRIP	0	3.03E-20	0.029760781
pt10-EOC11	BSNDKIW	0	3.03E-20	0.029760781

Table A.12: Inferred clone sizes within all live, Lin$^-$ cells for pt10 (related to Chapter 3). As in Table A.6 but for pt10.

A.22 Single-cell genotyping for pt11 (related to Chapter 3)

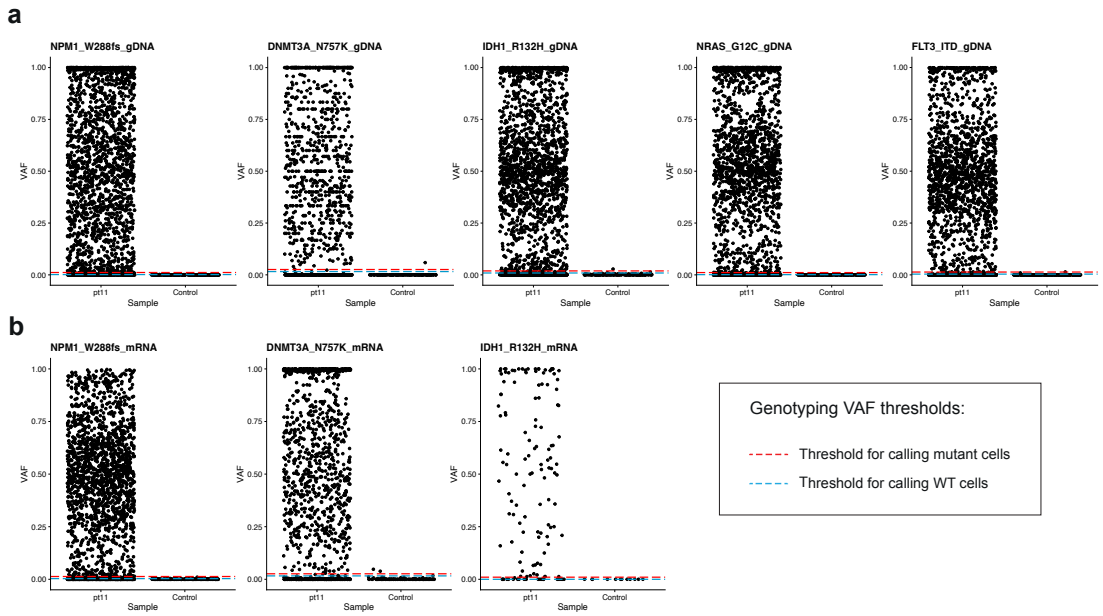


Figure A.6: Single-cell genotyping for pt11. (a) Single-cell variant allele frequency (scVAF) from gDNA amplicons across five mutant loci in pt11, comparing cells (dots) from pt11 and a WT control BM. Mutations are: *DNMT3A*-N757K, *NPM1*-W288fs*12, *IDH1*-R132H, *NRAS*-G12C, *FLT3*-ITD. Dotted lines indicate scVAF thresholds (section 2.8.2). Cells below the blue threshold are called WT. Cells above the red threshold are called mutant. Cells between the two threshold are called undetermined. Only cells passing single-cell genotyping QC are shown. **(b)** As in (a), but from cDNA amplicons.

A.23 Single-cell genotyping for pt15 (related to Chapter 3)

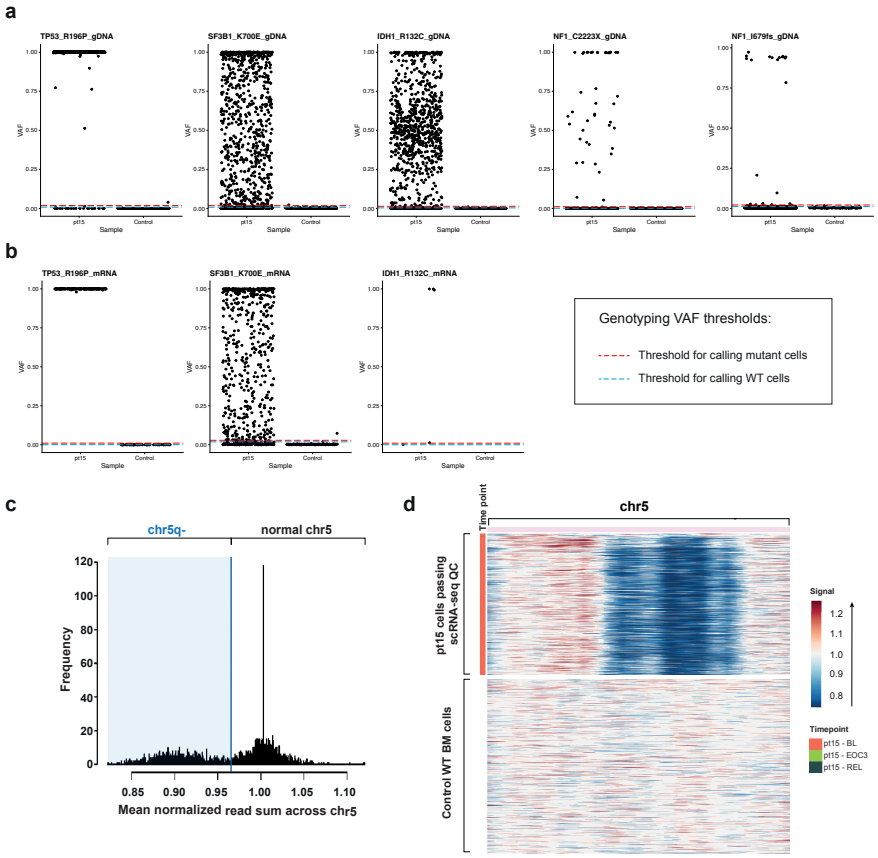


Figure A.7: Single-cell genotyping for pt15. (a) Same as in Figure A.6a but across 5 gDNA amplicons in pt15. Mutations are: *TP53*-R196P, *SF3B1*-K700E, *IDH1*-R132C, *NF1*-C2223*, *NF1*-I679fs. **(b)** As in (a), but from cDNA amplicons. **(c)** Single-cell CNA calling for pt15. del(5)(q11.2q33) was identified by karyotyping. Histogram of mean normalized read sum across chromosome 5 (the first 45% of genes across chromosome 5 were excluded, to avoid dilution of CNA signal) for each single cell (from pt15 or from a control BM). Cells within the blue shading are called chr5q- (i.e. they have del(5)(q11.2q33)). **(d)** inferCNV heatmaps for pt15 (top; n = 866 cells) and a healthy control BM (bottom; n = 1,057 cells), showing RNA expression signal across chromosome 5 (x-axis) in single cells (y-axis). The bar to the left indicates the time point for each cell. On the heatmap, red indicates higher and blue indicates lower signal. Cells passing scRNA-seq QC are shown.

A.24 Single-cell genotyping for pt20 (related to Chapter 3)

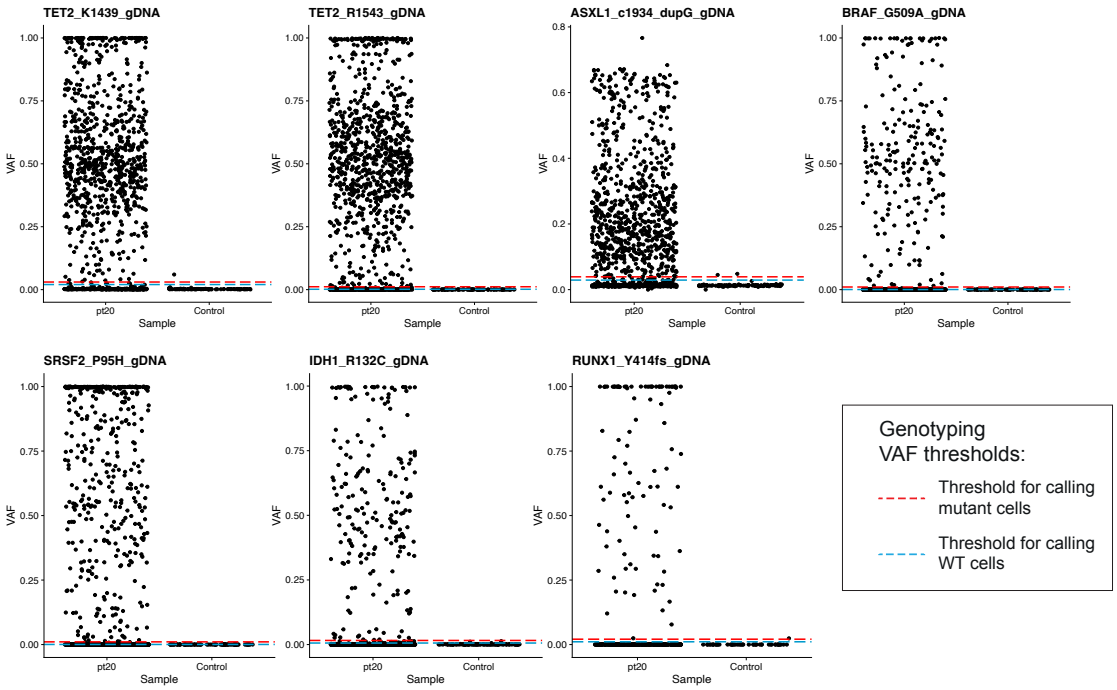


Figure A.8: Single-cell genotyping for pt20. Same as in Figure A.6a but across 7 gDNA amplicons in pt20. Mutations are: *TET2*-K1439fs*9 (*TET2*-K), *TET2*-R1543* (*TET2*-R), *ASXL1*-G646fs*11, *BRAF*-G469A, *SRSF2*-P95H, *IDH1*-R132C, and *RUNX1*-Y414fs*186.

A.25 Single-cell genotyping for pt5 (related to Chapter 3)

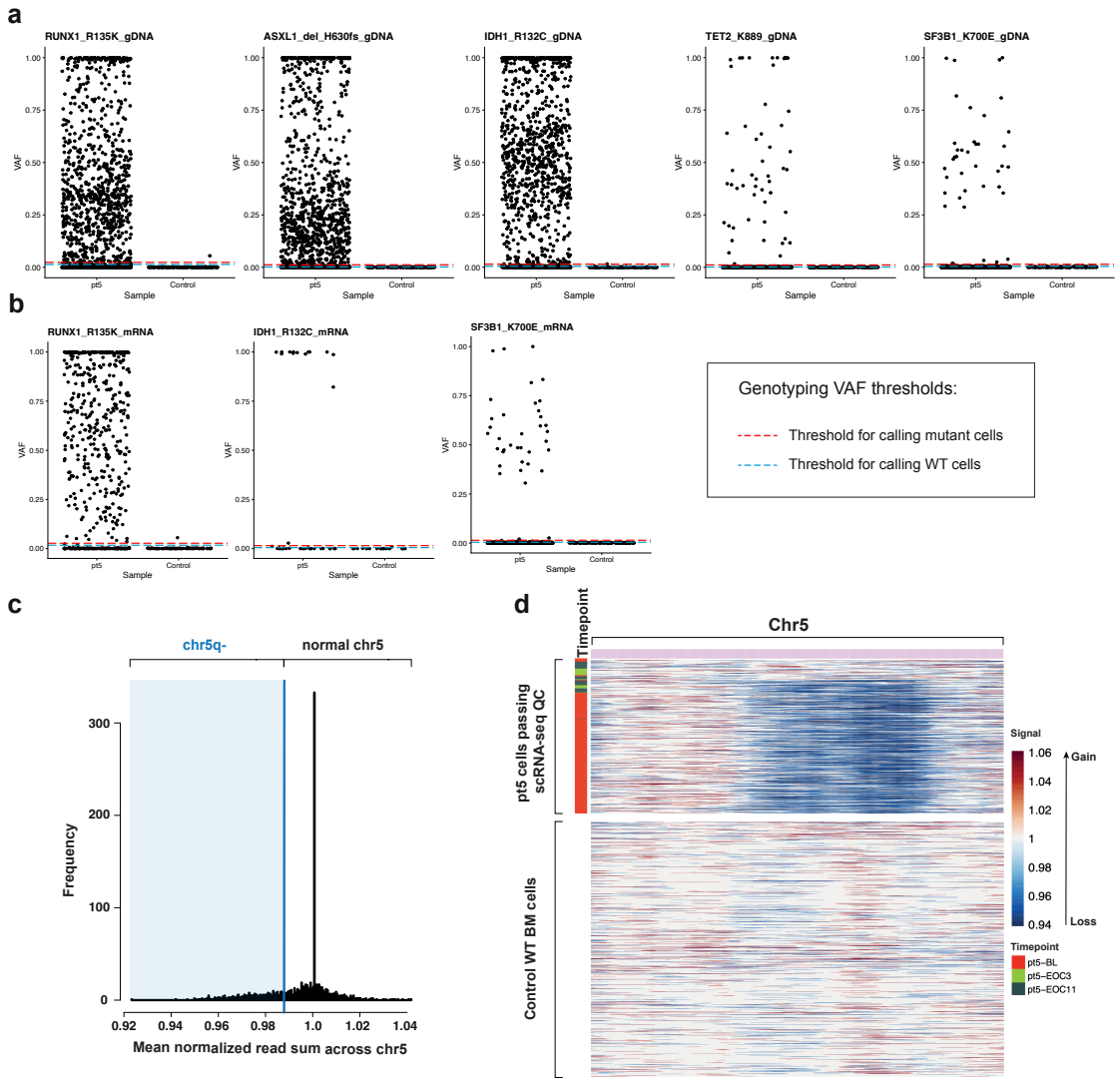


Figure A.9: Single-cell genotyping for pt5. (a) Same in Figure A.6a but across 5 gDNA amplicons in pt5. Mutations are: *RUNX1*-R162K, *ASXL1*-H630fs, *IDH1*-R132C, *TET2*-K889*, *SF3B1*-K700E. **(b)** As in (a), but from cDNA amplicons. **(c)** Same as in Figure A.7c but inferring del(5)(q22q35) in pt5. **(d)** Same as in Figure A.7d but for pt5 (top; n = 732 cells) and a healthy control BM (bottom; n = 1,197 cells).

A.26 Single-cell genotyping for pt14 (related to Chapter 3)

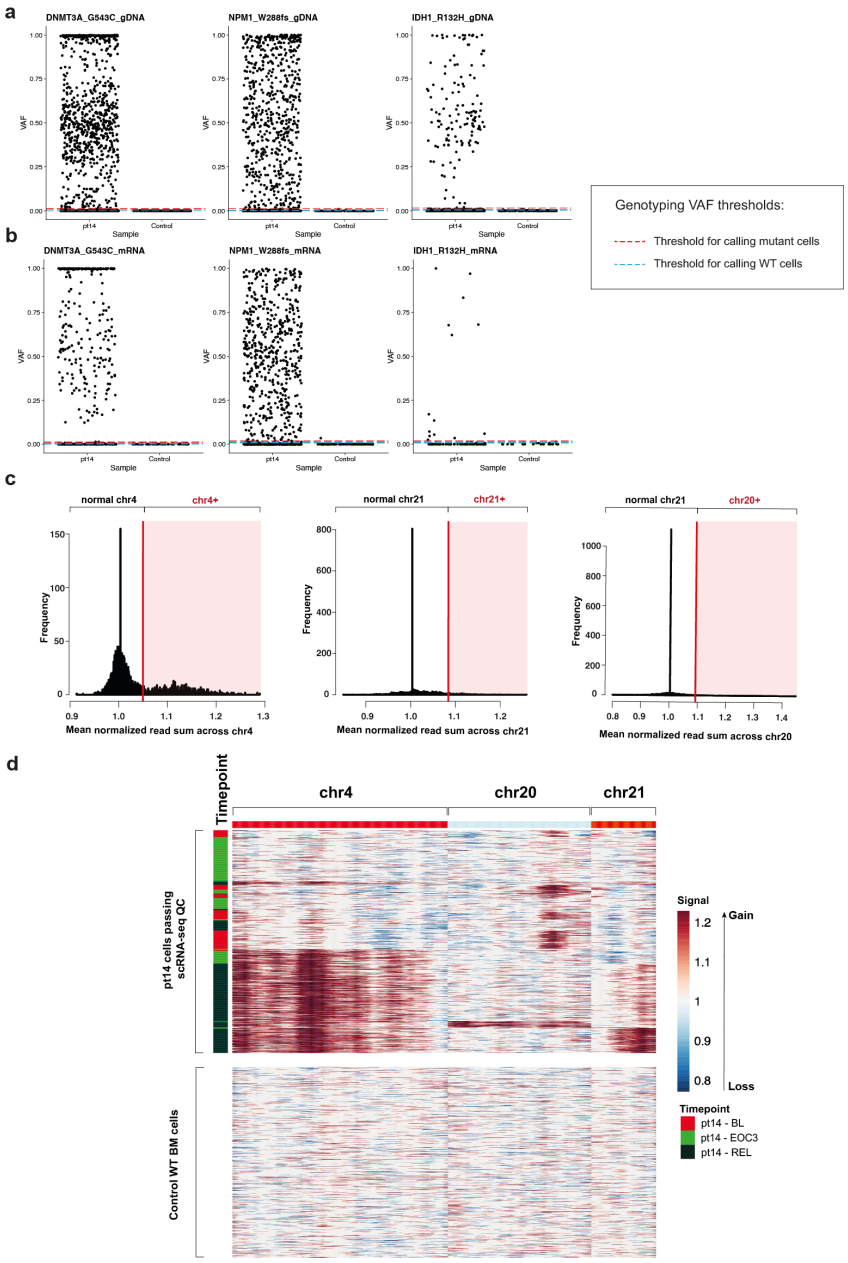


Figure A.10: Single-cell genotyping for pt14. **(a)** Same as in Figure A.6a but across 3 gDNA amplicons in pt14. Mutations are: *DNMT3A*-G543C, *NPM1*-W288fs*12, *IDH1*-R132H. **(b)** As in (a), but from cDNA amplicons. **(c)** Same as in Figure A.7c but inferring trisomy 4, 21, and 20 in pt14. Cells within the red shading are defined as harboring the CNA. **(d)** Same as in Figure A.7d but for pt14 (top; n = 1,251 cells) and a healthy control BM (bottom; n = 1,057 cells), showing chromosomes 4, 20, and 21.

A.27 Single-cell genotyping for pt9 (related to Chapter 3)

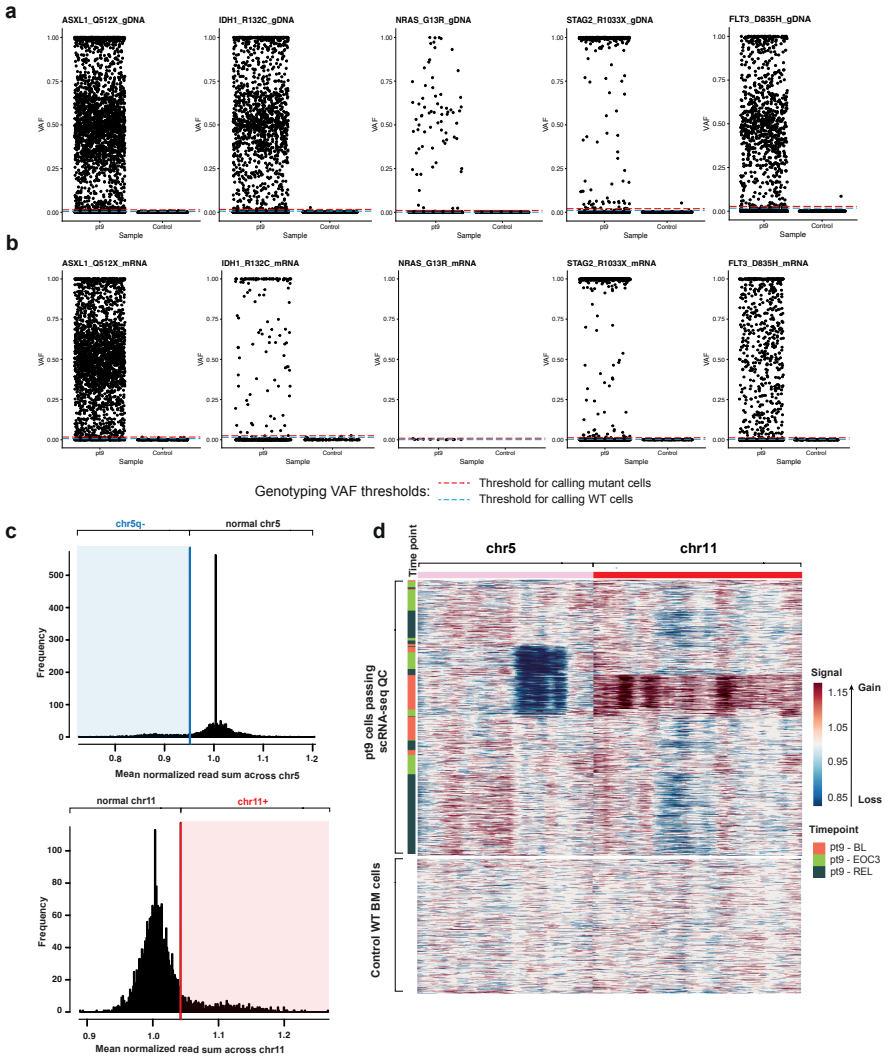


Figure A.11: Single-cell genotyping for pt9. (a) Same as in Figure A.6a but across 5 gDNA amplicons in pt9. Mutations are: *ASXL1*-Q512*, *NRAS*-G13R, *IDH1*-R132C, *STAG2*-R1033*, *FLT3*-D835H. **(b)** As in (a), but from cDNA amplicons. **(c)** Same as in Figure A.7c but inferring del(5)(q22q33) and trisomy 11. **(d)** Same as in Figure A.7d but for pt9 (top; n = 2,340 cells) and a healthy control BM (bottom; n = 1,057 cells), showing chromosomes 5 and 11.

A.28 Single-cell genotyping for pt4 (related to Chapter 3)

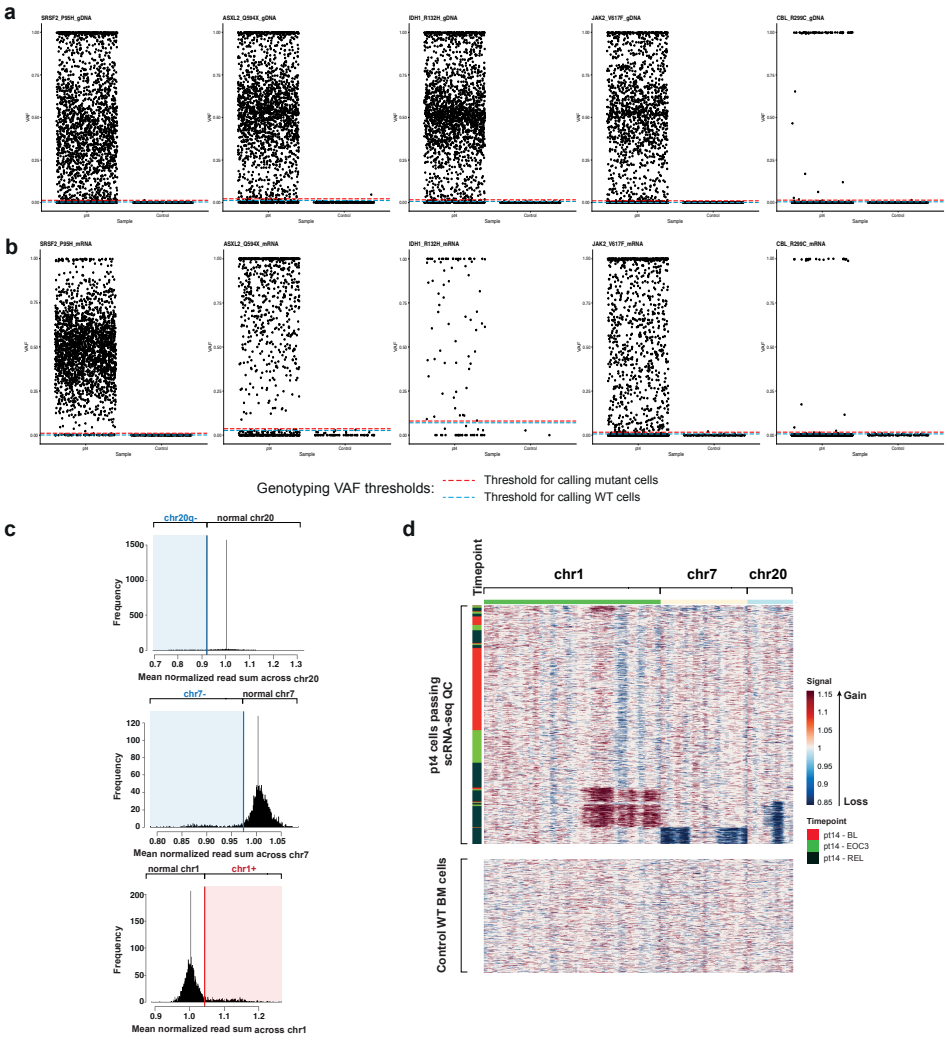


Figure A.12: Single-cell genotyping for pt4. **(a)** Same as in Figure A.6a but across 5 gDNA amplicons in pt4. Mutations are: *SRSF2*-P95H, *ASXL2*-Q594*, *IDH1*-R132H, *JAK2*-V617F, *CBL*-R299C **(b)** As in (a), but from cDNA amplicons. **(c)** Same as in Figure A.7c but inferring del(20)(q11.2q13.3), monosomy 7, and trisomy 1. For chromosome 1, the last 50% of genes across chromosome 1 were used. **(d)** Same as in Figure A.7d but for pt4 (top; n = 2,233 cells) and a healthy control BM (bottom; n = 1,057 cells), showing chromosomes 1, 7, and 20.

A.29 Single-cell genotyping for pt18 (related to Chapter 3)

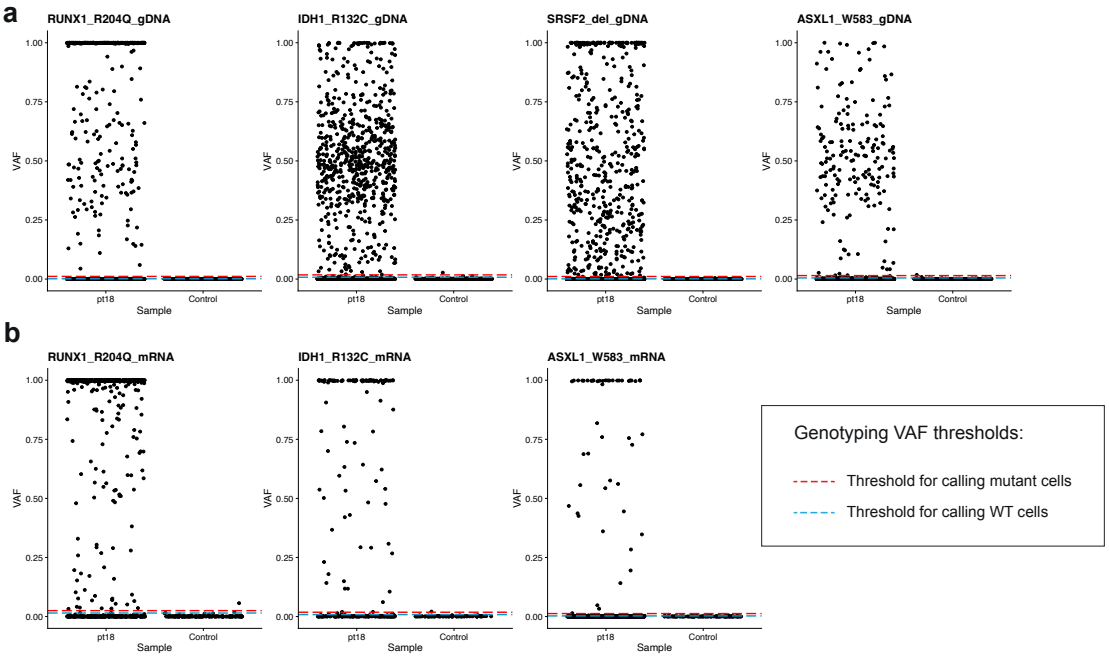


Figure A.13: Single-cell genotyping for pt18. **(a)** Same as in Figure A.6a but across 4 gDNA amplicons in pt18. Mutations are: *RUNX1*-R204Q, *IDH1*-R132H, *SRSF2*-P95_R102del, *ASXL1*-W583*. **(b)** As in (a), but from cDNA amplicons.

A.30 Single-cell genotyping for pt10 (related to Chapter 3)

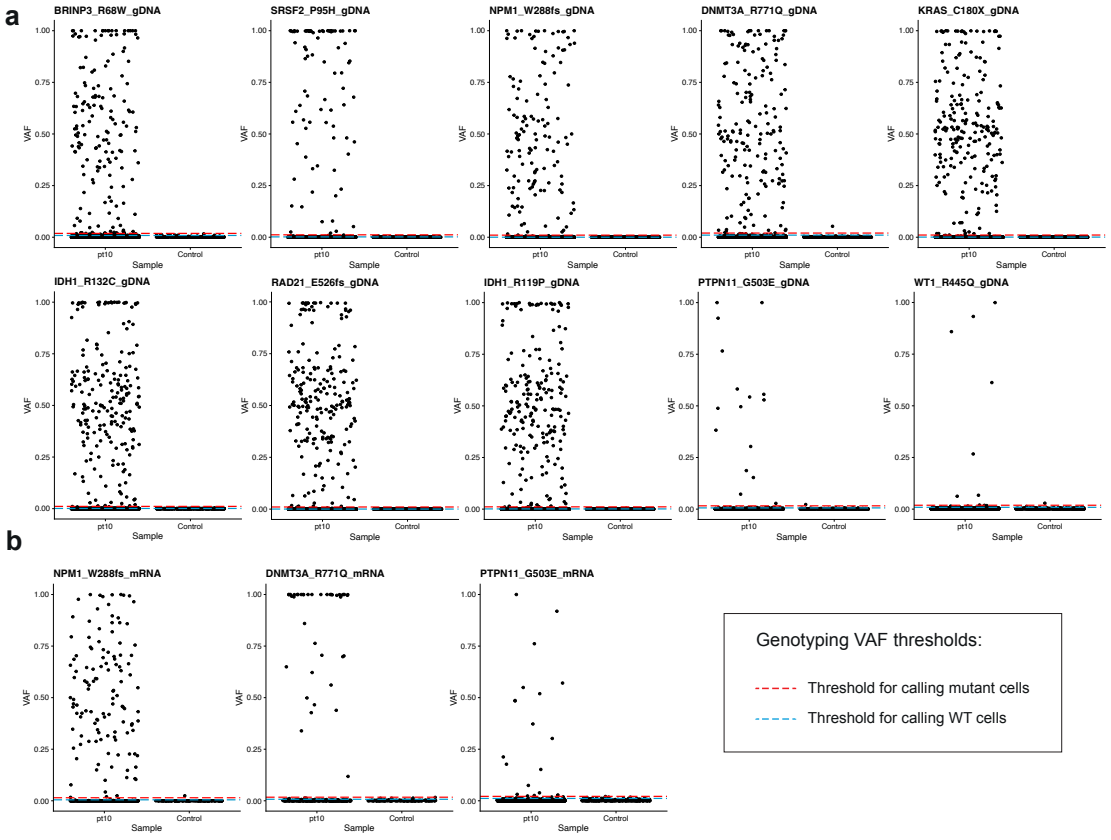


Figure A.14: Single-cell genotyping for pt10. (a) Same as in Figure A.6a but across 10 gDNA amplicons in pt10. Mutations are: *BRINP3*-R68W, *SRSF2*-P95H, *NPM1*-W288fs*12, *DNMT3A*-R771Q, *KRAS*-C180X, *IDH1*-R132S, *RAD21*-E526fs*9, *IDH1*-R119P, *PTPN11*-G503E, and *WT1*-R462Q. **(b)** As in (a), but from cDNA amplicons.

A.31 Genotyping primers used for GTAC (related to Chapters 4 and 5)

Locus	Primer pair	Forward primer	Reverse primer	Sample
ASXL1 exon 13	Pre-amplification	ACTTTCACGTATCAAACCACCC	CGGGCTTTAATGTCTGCGAG	Cell lines
CBL exon 15	Pre-amplification	CCAACACTGGTCCCGAGGA	GGATCACCATCCAGAGACAACC	Cell lines
HOXA9 exon 1	Pre-amplification	TGCTTGTGGTTCTCCTCCAG	CCAACTCGAGAGGCTTCCAG	Cell lines
NOTCH1 exon 6	Pre-amplification	GTGCCAGCTGATGCCAAATG	GGGACACTCGCAGTAGAAGG	Cell lines
NRAS exon 3	Pre-amplification	CCACACCCCCAGGATTCTTAC	TGCTCCTAGTACCTGTAGAGG	Cell lines
RUNX1 exon 4	Pre-amplification	GGGATGTTCCAGATGGCACT	TGGGTTTGTTGCCATGAAACG	Cell lines
TET2 exon 11	Pre-amplification	AACCAAAATATGCAGGGAGATGG	GGATTGCTCAGATTGGGTGG	Cell lines
DNMT3A exon 8	Pre-amplification	GCTGGTGTGGGGGAAACTG	AACCACATGACCCAGCGG	Cell lines
DNMT3A exon 23	Pre-amplification	CCCTCTCTGCCTTTTCTCCC	TGTCGCTACCTCAGTTTGCC	Cell lines
FLT3 exon 14	Pre-amplification	CAGCTACAGATGGTACAGGTGA	GACAAATGGTGAGTACGTGCATTT	Cell lines
NPM1 exon 11	Pre-amplification	TGTCTATGAAGTGTTGTGGTTCC	ACCACCGCTACTACTATGTCC	Cell lines
ASXL1 exon 13	Pre-amplification	GATGGTGGTGAGGCCTGTG	GCAGGTCCTCTCTCCTAGCT	AML
RUNX1 exon 5	Pre-amplification	TCTCTATCGTGTCCCCACAGG	CCCTGGGGGAAAGGTTGAAC	AML
TET2 exon 6	Pre-amplification	TGCCCTTATCTGCTGCAAGTG	GATTCCTTCCCACACCAGGAT	AML
BCOR exon 4	Pre-amplification	CCACATGAAAAAGATGGCTCCC	CCTTGTTTTTCAGCGACATGC	AML
BCOR exon 4	Pre-amplification	GCAGAGAGTGTTGGCCAGAG	GCTCCAGGAATGGAGTCGAG	AML
ASXL2 exon 10	Pre-amplification	GTCTAAGCAGCCCCACTGAAG	GCTCATTCAATGGTATCTCAGCC	AML
SH2B3 exon 6	Pre-amplification	GCCCCATCTCCAGAGTGAAA	AGGACCCTGAACTCCAGTGG	AML
ASXL1 exon 13	Nested (PCR1)	CCCTGGGTGGTTAAAGGTCAG	CAGGAGGACTCCGTGGTGG	Cell lines
CBL exon 15	Nested (PCR1)	TGGGTATGATGTCCCAAAGCC	AACCAGCCAAAGGAGGAGCT	Cell lines
HOXA9 exon 1	Nested (PCR1)	CGAAGGCGCCTTCTCTGAAA	CGGCGGCGGATTTGAAG	Cell lines
NOTCH1 exon 6	Nested (PCR1)	ACGGTGGCTACAACTGCG	GGCACAGTCATCAATGTTCTCG	Cell lines
NRAS exon 3	Nested (PCR1)	ACCTGTTTGTTGGACATACTGGA	TACACAGAGGAAGCCTTCGC	Cell lines
RUNX1 exon 4	Nested (PCR1)	TCACTGTGATGGCTGGCAAT	TCCACTTCGACCGACAAACC	Cell lines
TET2 exon 11	Nested (PCR1)	AGATGGTTTCAGCAGTTGTACCA	TAGAGGTGGCTCCCATGAAGT	Cell lines
DNMT3A exon 8	Nested (PCR1)	CTTCTCCTGGTGGCCAGG	GGGTGCCTTCAGCTGCTC	Cell lines
DNMT3A exon 23	Nested (PCR1)	TCTCCCCCAGGGTATTTGGT	TTTGCCCCCATGTCCCTTAC	Cell lines
FLT3 exon 14	Nested (PCR1)	ACCGGCTCCTCAGATAATGAGTA	CTAAATTTTCTCTTGGAAACTCCCA	Cell lines
NPM1 exon 11	Nested (PCR1)	TGAAGTGTTGTGGTTCCTTAACC	AGCAAAGGGTGGAGTTCCAT	Cell lines
ASXL1 exon 13	Nested (PCR1)	CGAGCACCCCTGGAAAGT	CAGTTCCGGCCTGGGTATG	AML
RUNX1 exon 5	Nested (PCR1)	CTTCACAAACCCACCGCAAGT	GCCCCAAGTGGATGCACTTAC	AML
TET2 exon 6	Nested (PCR1)	GGTGTTTGGGATGGAATGGTGA	CTCGCTCCCGCACCAAAC	AML
BCOR exon 4	Nested (PCR1)	TCCTCCAGGAAATGGTTGTGC	CTTCGTTAGGACTTGGCCCG	AML
BCOR exon 4	Nested (PCR1)	ACCGTGATTTCATCGCCCT	GTCCTTGCTTACGGTGAAGAC	AML
ASXL2 exon 10	Nested (PCR1)	CACAGTGAAAGAGGAGCCCC	CCTGCTGTGAGGTCCTGAAC	AML
SH2B3 exon 6	Nested (PCR1)	CTGATGCTCATGGAGTGTTCCT	ACCCCATACCTTGGCTATCCC	AML

Table A.13: Genotyping primers used for GTAC (related to Chapters 4 and 5). Note that, for the nested genotyping primers, only the locus-specific sequence is listed (without CS1 or CS2 adapters).

References

[1] Shilpa M. Hattangadi *et al.* "From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications". In: *Blood* 118.24 (Dec. 2011), pp. 6258–6268. URL: https://doi.org/10.1182/blood-2011-07-356006.

[2] Ron Sender and Ron Milo. "The distribution of cellular turnover in the human body". In: *Nature Medicine* 27 (1 2021), pp. 45–48. URL: https://doi.org/10.1038/s41591-020-01182-9.

[3] Barry Cooper. "The Origins of Bone Marrow as the Seedbed of Our Blood: From Antiquity to the Time of Osler". In: *Baylor University Medical Center Proceedings* 24.2 (2011). PMID: 21566758, pp. 115–118. URL: https://doi.org/10.1080/08998280.2011.11928697.

[4] Henry Lee-Six *et al.* "Population dynamics of normal human blood inferred from somatic mutations". In: *Nature* 561 (7724 2018), pp. 473–478. URL: https://doi.org/10.1038/s41586-018-0497-0.

[5] C M Baum *et al.* "Isolation of a candidate human hematopoietic stem-cell population". In: *Proceedings of the National Academy of Sciences* 89 (7 1992), pp. 2804–2808. URL: https://www.pnas.org/doi/abs/10.1073/pnas.89.7.2804.

[6] Mickie Bhatia *et al.* "Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice". In: *Proceedings of the National Academy of Sciences* 94 (10 1997), pp. 5320–5325. URL: https://www.pnas.org/doi/abs/10.1073/pnas.94.10.5320.

[7] Anne Wilson *et al.* "Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair". In: *Cell* 135 (6 2008), pp. 1118–1129. URL: https://doi.org/10.1016/j.cell.2008.10.048.

[8] A J Becker, E A McCulloch, and J E Till. "Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells". In: *Nature* 197 (4866 1963), pp. 452–454. URL: https://doi.org/10.1038/197452a0.

[9] Gerald J Spangrude, Shelly Heimfeld, and Irving L Weissman. "Purification and Characterization of Mouse Hematopoietic Stem Cells". In: *Science* 241 (4861 1988), pp. 58–62. URL: https://www.science.org/doi/abs/10.1126/science.2898810.

[10] Suzanne Kamel-Reid and John E Dick. "Engraftment of Immune-Deficient Mice with Human Hematopoietic Stem Cells". In: *Science* 242 (4886 1988), pp. 1706–1709. URL: https://www.science.org/doi/abs/10.1126/science.2904703.

[11] Sean J Morrison and Irving L Weissman. "The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype". In: *Immunity* 1 (8 1994), pp. 661–673. URL: https://www.sciencedirect.com/science/article/pii/107476139490037X.

[12] Faiyaz Notta *et al.* "Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment". In: *Science* 333 (6039 2011), pp. 218–221. URL: https://www.science.org/doi/abs/10.1126/science.1201219.

[13] Samuel H Cheshier *et al.* "In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells". In: *Proceedings of the National Academy of Sciences* 96 (6 1999), pp. 3120–3125. URL: https://www.pnas.org/doi/abs/10.1073/pnas.96.6.3120.

[14] Anne Wilson *et al.* "c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation". In: *Genes Development* 18 (22 2004), pp. 2747–2763. URL: http://genesdev.cshlp.org/content/18/22/2747.abstract.

[15] Elisa Laurenti *et al.* "CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells". In: *Cell Stem Cell* 16 (3 2015), pp. 302–313. URL: https://doi.org/10.1016/j.stem.2015.01.017.

[16] Peng Zou *et al.* "p57(Kip2) and p27(Kip1) Cooperate to Maintain Hematopoietic Stem Cell Quiescence through Interactions with Hsc70". In: *Cell Stem Cell* 9 (3 2011), pp. 247–261. URL: https://doi.org/10.1016/j.stem.2011.07.003.

[17] Nina Cabezas-Wallscheid *et al.* "Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy". In: *Cell* 169 (5 2017), 807–823.e19. URL: https://doi.org/10.1016/j.cell.2017.04.018.

[18] Robert A J Signer *et al.* "Haematopoietic stem cells require a highly regulated protein synthesis rate". In: *Nature* 509 (7498 2014), pp. 49–54. URL: https://doi.org/10.1038/nature13035.

[19] Theodore T Ho *et al.* "Autophagy maintains the metabolism and function of young and old stem cells". In: *Nature* 543 (7644 2017), pp. 205–210. URL: https://doi.org/10.1038/nature21388.

[20] Kyoko Ito *et al.* "Self-renewal of a purified Tie2+ hematopoietic stem cell population relies on mitochondrial clearance". In: *Science* 354 (6316 2016), pp. 1156–1160. URL: https://www.science.org/doi/abs/10.1126/science.aaf5530.

[21] Raymond Liang *et al.* "Restraining Lysosomal Activity Preserves Hematopoietic Stem Cell Quiescence and Potency". In: *Cell Stem Cell* 26 (3 2020), 359–376.e7. URL: https://doi.org/10.1016/j.stem.2020.01.013.

[22] Laura García-Prat *et al.* "TFEB-mediated endolysosomal activity controls human hematopoietic stem cell fate". In: *Cell Stem Cell* 28 (10 2021), 1838–1850.e10. URL: https://doi.org/10.1016/j.stem.2021.07.003.

[23] Dirk Loeffler *et al.* "Asymmetric lysosome inheritance predicts activation of haematopoietic stem cells". In: *Nature* 573 (7774 2019), pp. 426–429. URL: https://doi.org/10.1038/s41586-019-1531-6.

[24] Hanno Hock *et al.* "Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival". In: *Genes Development* 18 (19 2004), pp. 2336–2341. URL: http://genesdev.cshlp.org/content/18/19/2336.abstract.

[25] Stephen J Loughran *et al.* "The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells". In: *Nature Immunology* 9 (7 2008), pp. 810–819. URL: https://doi.org/10.1038/ni.1617.

[26] Elizabeth A Kruse *et al.* "Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage". In: *Proceedings of the National Academy of Sciences* 106 (33 2009), pp. 13814–13819. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0906556106.

[27] Hiromi Iwasaki *et al.* "Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation". In: *Blood* 106 (5 2005), pp. 1590–1600. URL: https://doi.org/10.1182/blood-2005-03-0860.

[28] Keisuke Kataoka *et al.* "Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity". In: *Journal of Experimental Medicine* 208 (12 2011), pp. 2403–2416. URL: https://doi.org/10.1084/jem.20110447.

[29] Susumu Goyama *et al.* "Evi-1 Is a Critical Regulator for Hematopoietic Stem Cells and Transformed Leukemic Cells". In: *Cell Stem Cell* 3 (2 2008), pp. 207–220. URL: https://doi.org/10.1016/j.stem.2008.06.002.

[30] Richard A Voit *et al.* "A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia". In: *Nature Immunology* 24 (1 2023), pp. 69–83. URL: https://doi.org/10.1038/s41590-022-01370-4.

[31] Fong-Ying Tsai *et al.* "An early haematopoietic defect in mice lacking the transcription factor GATA-2". In: *Nature* 371 (6494 1994), pp. 221–226. URL: https://doi.org/10.1038/371221a0.

[32] Alex J Tipping *et al.* "High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle". In: *Blood* 113 (12 2009), pp. 2661–2672. URL: https://doi.org/10.1182/blood-2008-06-161117.

[33] Karolina Komorowska *et al.* "Hepatic Leukemia Factor Maintains Quiescence of Hematopoietic Stem Cells and Protects the Stem Cell Pool during Regeneration". In: *Cell Reports* 21 (12 2017), pp. 3514–3523. URL: https://doi.org/10.1016/j.celrep.2017.11.084.

[34] Hui Zeng *et al.* "Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells". In: *The EMBO Journal* 23 (20 2004), pp. 4116–4125. URL: https://www.embopress.org/doi/abs/10.1038/sj.emboj.7600419.

[35] Hanno Hock *et al.* "Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells". In: *Nature* 431 (7011 2004), pp. 1002–1007. URL: https://doi.org/10.1038/nature02994.

[36] Olga Sirin *et al.* "The orphan nuclear receptor Nurr1 restricts the proliferation of haematopoietic stem cells". In: *Nature Cell Biology* 12 (12 2010), pp. 1213–1219. URL: https://doi.org/10.1038/ncb2125.

[37] Francesca Ficara *et al.* "Pbx1 Regulates Self-Renewal of Long-Term Hematopoietic Stem Cells by Maintaining Their Quiescence". In: *Cell Stem Cell* 2 (5 2008), pp. 484–496. URL: https://doi.org/10.1016/j.stem.2008.03.004.

[38] Yen K Lieu and E Premkumar Reddy. "Conditional c-myb knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation". In: *Proceedings of the National Academy of Sciences* 106 (51 2009), pp. 21689–21694. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0907623106.

[39] Yuko Kato *et al.* "Selective activation of STAT5 unveils its role in stem cell self-renewal in normal and leukemic hematopoiesis". In: *Journal of Experimental Medicine* 202 (1 2005), pp. 169–179. URL: https://doi.org/10.1084/jem.20042541.

[40] Sarah J Stein and Albert S Baldwin. "Deletion of the NF-κB subunit p65/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function". In: *Blood* 121 (25 2013), pp. 5015–5024. URL: https://doi.org/10.1182/blood-2013-02-486142.

[41] Chen Zhao *et al.* "Noncanonical NF-κB Signaling Regulates Hematopoietic Stem Cell Self-Renewal and Microenvironment Interactions". In: *Stem Cells* 30 (4 2012), pp. 709–718. URL: https://doi.org/10.1002/stem.1050.

[42] Jiang Zhu *et al.* "NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal". In: *Proceedings of the National Academy of Sciences* 102 (33 2005), pp. 11728–11733. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0503405102.

[43] Gerd Bungartz *et al.* "NF-Y is necessary for hematopoietic stem cell proliferation and survival". In: *Blood* 119 (6 2012), pp. 1380–1389. URL: https://doi.org/10.1182/blood-2011-06-359406.

[44] Zuzana Tothova *et al.* "FoxOs Are Critical Mediators of Hematopoietic Stem Cell Resistance to Physiologic Oxidative Stress". In: *Cell* 128 (2 2007), pp. 325–339. URL: https://doi.org/10.1016/j.cell.2007.01.003.

[45] Kana Miyamoto *et al.* "Foxo3a Is Essential for Maintenance of the Hematopoietic Stem Cell Pool". In: *Cell Stem Cell* 1 (1 2007), pp. 101–112. URL: https://www.sciencedirect.com/science/article/pii/S1934590907000021.

[46] H Jeffrey Lawrence *et al.* "Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells". In: *Blood* 106 (12 2005), pp. 3988–3994. URL: https://doi.org/10.1182/blood-2005-05-2003.

[47] Jennifer Antonchuk, Guy Sauvageau, and R Keith Humphries. "HOXB4-Induced Expansion of Adult Hematopoietic Stem Cells Ex Vivo". In: *Cell* 109 (1 2002), pp. 39–45. URL: https://doi.org/10.1016/S0092-8674(02)00697-9.

[48] Neal A Fischbach *et al.* "HOXB6 overexpression in murine bone marrow immortalizes a myelomonocytic precursor in vitro and causes hematopoietic stem cell expansion and acute myeloid leukemia in vivo". In: *Blood* 105 (4 2005), pp. 1456–1466. URL: https://doi.org/10.1182/blood-2004-04-1583.

[49] Shinya Kimura *et al.* "Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin". In: *Proceedings of the National Academy of Sciences* 95 (3 1998), pp. 1195–1200. URL: https://www.pnas.org/doi/abs/10.1073/pnas.95.3.1195.

[50] Charlotte E Edling and Bengt Hallberg. "c-Kit—A hematopoietic cell essential receptor tyrosine kinase". In: *The International Journal of Biochemistry Cell Biology* 39 (11 2007), pp. 1995–1998. URL: https://www.sciencedirect.com/science/article/pii/S1357272507000167.

[51] Tannishtha Reya *et al.* "A role for Wnt signalling in self-renewal of haematopoietic stem cells". In: *Nature* 423 (6938 2003), pp. 409–414. URL: https://doi.org/10.1038/nature01593.

[52] Vincenzo Calvanese *et al.* "MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment". In: *Nature* 576 (7786 2019), pp. 281–286. URL: https://doi.org/10.1038/s41586-019-1790-2.

[53] Craig D Jude *et al.* "Unique and Independent Roles for MLL in Adult Hematopoietic Stem Cells and Progenitors". In: *Cell Stem Cell* 1 (3 2007), pp. 324–337. URL: https://doi.org/10.1016/j.stem.2007.05.019.

[54] Kathryn A McMahon *et al.* "Mll Has a Critical Role in Fetal and Adult Hematopoietic Stem Cell Self-Renewal". In: *Cell Stem Cell* 1 (3 2007), pp. 338–345. URL: https://doi.org/10.1016/j.stem.2007.07.002.

[55] Thomas A Milne *et al.* "MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters". In: *Molecular Cell* 10 (5 2002), pp. 1107–1117. URL: https://doi.org/10.1016/S1097-2765(02)00741-4.

[56] In-kyung Park *et al.* "Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells". In: *Nature* 423 (6937 2003), pp. 302–305. URL: https://doi.org/10.1038/nature01587.

[57] Atsushi Iwama *et al.* "Enhanced Self-Renewal of Hematopoietic Stem Cells Mediated by the Polycomb Gene Product Bmi-1". In: *Immunity* 21 (6 2004), pp. 843–851. URL: https://doi.org/10.1016/j.immuni.2004.11.004.

[58] Huafeng Xie *et al.* "Polycomb Repressive Complex 2 Regulates Normal Hematopoietic Stem Cell Function in a Developmental-Stage-Specific Manner". In: *Cell Stem Cell* 14 (1 2014), pp. 68–80. URL: https://www.sciencedirect.com/science/article/pii/S1934590913004475.

[59] Leonie M Kamminga *et al.* "The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion". In: *Blood* 107 (5 2006), pp. 2170–2179. URL: https://doi.org/10.1182/blood-2005-09-3585.

[60] Ian J Majewski *et al.* "Polycomb Repressive Complex 2 (PRC2) Restricts Hematopoietic Stem Cell Activity". In: *PLOS Biology* 6 (4 2008), e93. URL: https://doi.org/10.1371/journal.pbio.0060093.

[61] Julie Lessard *et al.* "Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation". In: *Genes Development* 13 (20 1999), pp. 2691–2703. URL: http://genesdev.cshlp.org/content/13/20/2691.abstract.

[62] Grant A Challen *et al.* "Dnmt3a is essential for hematopoietic stem cell differentiation". In: *Nature Genetics* 44 (1 2012), pp. 23–31. URL: https://doi.org/10.1038/ng.1009.

[63] Mira Jeong *et al.* "Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo". In: *Cell Reports* 23 (1 2018), pp. 1–10. URL: https://doi.org/10.1016/j.celrep.2018.03.025.

[64] Elizabeth L Ostrander *et al.* "Divergent Effects of Dnmt3a and Tet2 Mutations on Hematopoietic Progenitor Cell Fitness". In: *Stem Cell Reports* 14 (4 2020), pp. 551–560. URL: https://doi.org/10.1016/j.stemcr.2020.02.011.

[65] Kelly Moran-Crusio *et al.* "Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation". In: *Cancer Cell* 20 (1 2011), pp. 11–24. URL: https://doi.org/10.1016/j.ccr.2011.06.001.

[66] Cyril Quivoron *et al.* "TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis". In: *Cancer Cell* 20 (1 2011), pp. 25–38. URL: https://doi.org/10.1016/j.ccr.2011.06.003.

[67] Naoya Takayama *et al.* "The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization". In: *Cell Stem Cell* 28 (3 2021), 488–501.e10. URL: https://doi.org/10.1016/j.stem.2020.11.001.

[68] Brad Dykstra *et al.* "Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells". In: *Journal of Experimental Medicine* 208 (13 2011), pp. 2691–2703. URL: https://doi.org/10.1084/jem.20111490.

[69] Grant A Challen *et al.* "Distinct Hematopoietic Stem Cell Subtypes Are Differentially Regulated by TGF-beta1". In: *Cell Stem Cell* 6 (3 2010), pp. 265–278. URL: https://doi.org/10.1016/j.stem.2010.02.002.

[70] Ryo Yamamoto *et al.* "Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells". In: *Cell* 154 (5 2013), pp. 1112–1126. URL: https://doi.org/10.1016/j.cell.2013.08.007.

[71] Alejandra Sanjuan-Pla *et al.* "Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy". In: *Nature* 502 (7470 2013), pp. 232–236. URL: https://doi.org/10.1038/nature12495.

[72] Joana Carrelha *et al.* "Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells". In: *Nature* 554 (7690 2018), pp. 106–111. URL: https://doi.org/10.1038/nature25455.

[73] Joana Carrelha *et al.* "Alternative platelet differentiation pathways initiated by nonhierarchically related hematopoietic stem cells". In: *Nature Immunology* 25 (6 2024), pp. 1007–1019. URL: https://doi.org/10.1038/s41590-024-01845-6.

[74] Alejo E Rodriguez-Fraticelli *et al.* "Clonal analysis of lineage fate in native haematopoiesis". In: *Nature* 553 (7687 2018), pp. 212–216. URL: https://doi.org/10.1038/nature25168.

[75] Merve Aksöz *et al.* "Hematopoietic stem cell heterogeneity and age-associated platelet bias are evolutionarily conserved". In: *Science Immunology* 9 (98 2024), eadk3469. URL: https://www.science.org/doi/abs/10.1126/sciimmunol.adk3469.

[76] Yiran Meng *et al.* "Epigenetic programming defines haematopoietic stem cell fate restriction". In: *Nature Cell Biology* 25 (6 2023), pp. 812–822. URL: https://doi.org/10.1038/s41556-023-01137-5.

[77] Emily Mitchell *et al.* "Clonal dynamics of haematopoiesis across the human lifespan". In: *Nature* 606 (7913 2022), pp. 343–350. URL: https://doi.org/10.1038/s41586-022-04786-y.

[78] Wendy W Pang *et al.* "Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age". In: *Proceedings of the National Academy of Sciences* 108 (50 2011), pp. 20012–20017. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1116110108.

[79] Andy G X Zeng *et al.* "A hematopoietic stem cell subset that retains memory of prior inflammatory stress accumulates in aging and clonal hematopoiesis". In: *bioRxiv* (2023), p. 2023.09.11.557271. URL: https://www.biorxiv.org/content/biorxiv/early/2023/09/13/2023.09.11.557271.full.pdf.

[80] Niels Asger Jakobsen *et al.* "Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging". In: *Cell Stem Cell* 31 (8 2024), 1127–1144.e17. URL: https://doi.org/10.1016/j.stem.2024.05.010.

[81] Lars Velten *et al.* "Human haematopoietic stem cell lineage commitment is a continuous process". In: *Nature Cell Biology* 19 (4 2017), pp. 271–281. URL: https://doi.org/10.1038/ncb3493.

[82] Faiyaz Notta *et al.* "Distinct routes of lineage development reshape the human blood hierarchy across ontogeny". In: *Science* 351 (6269 2016), aab2116. URL: https://www.science.org/doi/abs/10.1126/science.aab2116.

[83] Franziska Paul *et al.* "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors". In: *Cell* 163 (7 2015), pp. 1663–1677. URL: https://doi.org/10.1016/j.cell.2015.11.013.

[84] Elisa Laurenti and Berthold Göttgens. "From haematopoietic stem cells to complex differentiation landscapes". In: *Nature* 553 (7689 2018), pp. 418–426. URL: https://doi.org/10.1038/nature25022.

[85] Ravindra Majeti, Christopher Y Park, and Irving L Weissman. "Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood". In: *Cell Stem Cell* 1 (6 2007), pp. 635–645. URL: https://doi.org/10.1016/j.stem.2007.10.001.

[86] Eric M Pietras *et al.* "Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions". In: *Cell Stem Cell* 17 (1 2015), pp. 35–46. URL: https://doi.org/10.1016/j.stem.2015.05.003.

[87] Nina Cabezas-Wallscheid *et al.* "Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis". In: *Cell Stem Cell* 15 (4 2014), pp. 507–522. URL: https://www.sciencedirect.com/science/article/pii/S1934590914003014.

[88] Jeffrey M Granja *et al.* "Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia". In: *Nature Biotechnology* 37 (12 2019), pp. 1458–1465. URL: https://doi.org/10.1038/s41587-019-0332-7.

[89] Roy Drissen *et al.* "Identification of two distinct pathways of human myelopoiesis". In: *Science Immunology* 4 (35 2019), eaau7148. URL: https://www.science.org/doi/abs/10.1126/sciimmunol.aau7148.

[90] Dimitris Karamitros *et al.* "Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells". In: *Nature Immunology* 19 (1 2018), pp. 85–97. URL: https://doi.org/10.1038/s41590-017-0001-2.

[91] Andy G X Zeng *et al.* "Precise single-cell transcriptomic mapping of normal and leukemic cell states reveals unconventional lineage priming in acute myeloid leukemia". In: *bioRxiv* (2023), p. 2023.12.26.573390. URL: https://www.biorxiv.org/content/biorxiv/early/2023/12/27/2023.12.26.573390.full.pdf.

[92] M Ryan Corces *et al.* "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution". In: *Nature Genetics* 48 (10 2016), pp. 1193–1203. URL: https://doi.org/10.1038/ng.3646.

[93] Yu Wei Zhang *et al.* "Hyaluronic acid–GPRC5C signalling promotes dormancy in haematopoietic stem cells". In: *Nature Cell Biology* 24 (7 2022), pp. 1038–1048. URL: https://doi.org/10.1038/s41556-022-00931-x.

[94] Kerstin B Kaufmann *et al.* "A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness". In: *Nature Immunology* 22 (6 2021), pp. 723–734. URL: https://doi.org/10.1038/s41590-021-00925-1.

[95] Hideyuki Oguro *et al.* "Poised Lineage Specification in Multipotential Hematopoietic Stem and Progenitor Cells by the Polycomb Protein Bmi1". In: *Cell Stem Cell* 6 (3 2010), pp. 279–286. URL: https://doi.org/10.1016/j.stem.2010.01.005.

[96] Anne Y Lai and Motonari Kondo. "Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors". In: *Journal of Experimental Medicine* 203 (8 2006), pp. 1867–1873. URL: https://doi.org/10.1084/jem.20060697.

[97] Leïla Perié *et al.* "The Branching Point in Erythro-Myeloid Differentiation". In: *Cell* 163 (7 2015), pp. 1655–1662. URL: https://doi.org/10.1016/j.cell.2015.11.059.

[98] Jörgen Adolfsson *et al.* "Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment". In: *Cell* 121 (2 2005), pp. 295–306. URL: https://www.sciencedirect.com/science/article/pii/S0092867405001583.

[99] Robert Månsson *et al.* "Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors". In: *Immunity* 26 (4 2007), pp. 407–419. URL: https://doi.org/10.1016/j.immuni.2007.02.013.

[100] Nicolas Goardon *et al.* "Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia". In: *Cancer Cell* 19 (1 2011), pp. 138–152. URL: https://www.sciencedirect.com/science/article/pii/S153561081000526X.

[101] Toshimi Yoshida *et al.* "Early hematopoietic lineage restrictions directed by Ikaros". In: *Nature Immunology* 7 (4 2006), pp. 382–391. URL: https://doi.org/10.1038/ni1314.

[102] Yojiro Arinobu *et al.* "Reciprocal Activation of GATA-1 and PU.1 Marks Initial Specification of Hematopoietic Stem Cells into Myeloerythroid and Myelolymphoid Lineages". In: *Cell Stem Cell* 1 (4 2007), pp. 416–427. URL: https://doi.org/10.1016/j.stem.2007.07.004.

[103] Yojiro Arinobu *et al.* "Developmental checkpoints of the basophil/mast cell lineages in adult murine hematopoiesis". In: *Proceedings of the National Academy of Sciences* 102 (50 2005), pp. 18105–18110. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0509148102.

[104] Shalin H Naik *et al.* "Diverse and heritable lineage imprinting of early haematopoietic progenitors". In: *Nature* 496 (7444 2013), pp. 229–232. URL: https://doi.org/10.1038/nature12013.

[105] Lisa A Kohn *et al.* "Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin". In: *Nature Immunology* 13 (10 2012), pp. 963–971. URL: https://doi.org/10.1038/ni.2405.

[106] Sergei Doulatov *et al.* "Induction of Multipotential Hematopoietic Progenitors from Human Pluripotent Stem Cells via Respecification of Lineage-Restricted Precursors". In: *Cell Stem Cell* 13 (4 2013), pp. 459–470. URL: https://doi.org/10.1016/j.stem.2013.09.002.

[107] Anne Galy *et al.* "Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset". In: *Immunity* 3 (4 1995), pp. 459–473. URL: https://www.sciencedirect.com/science/article/pii/1074761395901752.

[108] Cornelis J H Pronk *et al.* "Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy". In: *Cell Stem Cell* 1 (4 2007), pp. 428–442. URL: https://www.sciencedirect.com/science/article/pii/S1934590907000719.

[109] Jaeyop Lee *et al.* "Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow". In: *Journal of Experimental Medicine* 212 (3 2015), pp. 385–399. URL: https://doi.org/10.1084/jem.20141442.

[110] Lynn Quek *et al.* "Genetically distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage". In: *Journal of Experimental Medicine* 213 (8 2016), pp. 1513–1535. URL: https://doi.org/10.1084/jem.20151775.

[111] Stephen L Nutt and Barbara L Kee. "The Transcriptional Regulation of B Cell Lineage Commitment". In: *Immunity* 26 (6 2007), pp. 715–725. URL: https://doi.org/10.1016/j.immuni.2007.05.010.

[112] Stuart H Orkin and Leonard I Zon. "Hematopoiesis: An Evolving Paradigm for Stem Cell Biology". In: *Cell* 132 (4 2008), pp. 631–644. URL: https://doi.org/10.1016/j.cell.2008.01.025.

[113] Stephen L Nutt and Michaël Chopin. "Transcriptional Networks Driving Dendritic Cell Differentiation and Function". In: *Immunity* 52 (6 2020), pp. 942–956. URL: https://doi.org/10.1016/j.immuni.2020.05.005.

[114] Hiromi Iwasaki and Koichi Akashi. "Myeloid Lineage Commitment from the Hematopoietic Stem Cell". In: *Immunity* 26 (6 2007), pp. 726–740. URL: https://doi.org/10.1016/j.immuni.2007.06.004.

[115] S I Kim and E H Bresnick. "Transcriptional control of erythropoiesis: emerging mechanisms and principles". In: *Oncogene* 26 (47 2007), pp. 6777–6794. URL: https://doi.org/10.1038/sj.onc.1210761.

[116] Frank Rosenbauer and Daniel G Tenen. "Transcription factors in myeloid development: balancing differentiation with transformation". In: *Nature Reviews Immunology* 7 (2 2007), pp. 105–117. URL: https://doi.org/10.1038/nri2024.

[117] Edward W Scott *et al.* "Requirement of Transcription Factor PU.1 in the Development of Multiple Hematopoietic Lineages". In: *Science* 265 (5178 1994), pp. 1573–1577. URL: https://www.science.org/doi/abs/10.1126/science.8079170.

[118] Daniel G Tenen *et al.* "Transcription Factors, Normal Myeloid Development, and Leukemia". In: *Blood* 90 (2 1997), pp. 489–519. URL: https://www.sciencedirect.com/science/article/pii/S0006497120589234.

[119] Aleksandar Dakic *et al.* "PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis". In: *Journal of Experimental Medicine* 201 (9 2005), pp. 1487–1502. URL: https://doi.org/10.1084/jem.20050075.

[120] Claus Nerlov and Thomas Graf. "PU.1 induces myeloid lineage commitment in multipotent hematopoieticprogenitors". In: *Genes Development* 12 (15 Aug. 1998). 10.1101/gad.12.15.2403, pp. 2403–2412. URL: http://genesdev.cshlp.org/content/12/15/2403.abstract.

[121] Larysa Pevny *et al.* "Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1". In: *Nature* 349 (6306 1991), pp. 257–260. URL: https://doi.org/10.1038/349257a0.

[122] Jenna L Galloway *et al.* "Loss of Gata1 but Not Gata2 Converts Erythropoiesis to Myelopoiesis in Zebrafish Embryos". In: *Developmental Cell* 8 (1 2005), pp. 109–116. URL: https://doi.org/10.1016/j.devcel.2004.12.001.

[123] Paresh Vyas *et al.* "Consequences of GATA-1 Deficiency in Megakaryocytes and Platelets". In: *Blood* 93 (9 1999), pp. 2867–2875. URL: https://www.sciencedirect.com/science/article/pii/S0006497120596079.

[124] Ramesh A Shivdasani *et al.* "A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development". In: *The EMBO Journal* 16 (13 1997), pp. 3965–3973. URL: https://www.embopress.org/doi/abs/10.1093/emboj/16.13.3965.

[125] Channing Yu *et al.* "Targeted Deletion of a High-Affinity GATA-binding Site in the GATA-1 Promoter Leads to Selective Loss of the Eosinophil Lineage In Vivo". In: *Journal of Experimental Medicine* 195 (11 2002), pp. 1387–1395. URL: https://doi.org/10.1084/jem.20020656.

[126] Yuichiro Nei *et al.* "GATA-1 regulates the generation and function of basophils". In: *Proceedings of the National Academy of Sciences* 110 (46 2013), pp. 18620–18625. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1311668110.

[127] Anna Rita Migliaccio *et al.* "GATA-1 as a Regulator of Mast Cell Differentiation Revealed by the Phenotype of the GATA-1low Mouse Mutant". In: *Journal of Experimental Medicine* 197 (3 2003), pp. 281–296. URL: https://doi.org/10.1084/jem.20021149.

[128] Hiromi Iwasaki *et al.* "GATA-1 Converts Lymphoid and Myelomonocytic Progenitors into the Megakaryocyte/Erythrocyte Lineages". In: *Immunity* 19 (3 2003), pp. 451–462. URL: https://doi.org/10.1016/S1074-7613(03)00242-5.

[129] Natasha Rekhtman *et al.* "Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells". In: *Genes Development* 13 (11 1999), pp. 1398–1411. URL: http://genesdev.cshlp.org/content/13/11/1398.abstract.

[130] Claus Nerlov *et al.* "GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription". In: *Blood* 95 (8 2000), pp. 2543–2551. URL: https://www.sciencedirect.com/science/article/pii/S000649712064143X.

[131] Pu Zhang *et al.* "Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1". In: *Proceedings of the National Academy of Sciences* 96 (15 1999), pp. 8705–8710. URL: https://www.pnas.org/doi/abs/10.1073/pnas.96.15.8705.

[132] Pu Zhang *et al.* "PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding". In: *Blood* 96 (8 2000), pp. 2641–2648. URL: https://doi.org/10.1182/blood.V96.8.2641.

[133] Philipp S Hoppe *et al.* "Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios". In: *Nature* 535 (7611 2016), pp. 299–302. URL: https://doi.org/10.1038/nature18320.

[134] Dong-Er Zhang *et al.* "Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding proteinα-deficientmice". In: *Proceedings of the National Academy of Sciences* 94 (2 1997), pp. 569–574. URL: https://www.pnas.org/doi/abs/10.1073/pnas.94.2.569.

[135] Pu Zhang *et al.* "Enhancement of Hematopoietic Stem Cell Repopulating Capacity and Self-Renewal in the Absence of the Transcription Factor C/EBPα". In: *Immunity* 21 (6 2004), pp. 853–863. URL: https://www.sciencedirect.com/science/article/pii/S1074761304003450.

[136] Jörg Cammenga *et al.* "Induction of C/EBPα activity alters gene expression and differentiation of human CD34+ cells". In: *Blood* 101 (6 2003), pp. 2206–2214. URL: https://doi.org/10.1182/blood-2002-05-1546.

[137] Hein Schepers *et al.* "Reintroduction of C/EBPα in leukemic CD34+ stem/progenitor cells impairs self-renewal and partially restores myelopoiesis". In: *Blood* 110 (4 2007), pp. 1317–1325. URL: https://doi.org/10.1182/blood-2006-10-052175.

[138] Pu Zhang *et al.* "Upregulation of Interleukin 6 and Granulocyte Colony-Stimulating Factor Receptors by Transcription Factor CCAAT Enhancer Binding Protein α (C/EBPα) Is Critical for Granulopoiesis". In: *Journal of Experimental Medicine* 188 (6 1998), pp. 1173–1184. URL: https://doi.org/10.1084/jem.188.6.1173.

[139] Hyung Chan Suh *et al.* "C/EBPα determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation". In: *Blood* 107 (11 2006), pp. 4308–4316. URL: https://doi.org/10.1182/blood-2005-06-2216.

[140] Roberto Avellino and Ruud Delwel. "Expression and regulation of C/EBPα in normal myelopoiesis and in malignant transformation". In: *Blood* 129 (15 2017), pp. 2083–2091. URL: https://doi.org/10.1182/blood-2016-09-687822.

[141] Ryuya Yamanaka *et al.* "Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein -deficient mice". In: *Proceedings of the National Academy of Sciences* 94 (24 1997), pp. 13187–13192. URL: https://www.pnas.org/doi/abs/10.1073/pnas.94.24.13187.

[142] Malene Digmann Bjerregaard *et al.* "The in vivo profile of transcription factors during neutrophil differentiation in human bone marrow". In: *Blood* 101 (11 2003), pp. 4322–4332. URL: https://doi.org/10.1182/blood-2002-03-0835.

[143] Tomohiko Tamura *et al.* "ICSBP Directs Bipotential Myeloid Progenitor Cells to Differentiate into Mature Macrophages". In: *Immunity* 13 (2 2000), pp. 155–165. URL: https://doi.org/10.1016/S1074-7613(00)00016-9.

[144] Hanno Hock *et al.* "Intrinsic Requirement for Zinc Finger Transcription Factor Gfi-1 in Neutrophil Differentiation". In: *Immunity* 18 (1 2003), pp. 109–120. URL: https://doi.org/10.1016/S1074-7613(02)00501-0.

[145] Shane R Horman *et al.* "Gfi1 integrates progenitor versus granulocytic transcriptional programming". In: *Blood* 113 (22 2009), pp. 5466–5475. URL: https://doi.org/10.1182/blood-2008-09-179747.

[146] Marella de Bruijn and Elaine Dzierzak. "Runx transcription factors in the development and function of the definitive hematopoietic system". In: *Blood* 129 (15 2017), pp. 2061–2069. URL: https://doi.org/10.1182/blood-2016-12-689109.

[147] Trista E North *et al.* "Runx1 Is Expressed in Adult Mouse Hematopoietic Stem Cells and Differentiating Myeloid and Lymphoid Cells, But Not in Maturing Erythroid Cells". In: *Stem Cells* 22 (2 2004), pp. 158–168. URL: https://doi.org/10.1634/stemcells.22-2-158.

[148] Robert B Lorsbach *et al.* "Role of RUNX1 in adult hematopoiesis: analysis of RUNX1-IRES-GFP knock-in mice reveals differential lineage expression". In: *Blood* 103 (7 2004), pp. 2522–2529. URL: https://doi.org/10.1182/blood-2003-07-2439.

[149] Joseph D Growney *et al.* "Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype". In: *Blood* 106 (2 2005), pp. 494–504. URL: https://doi.org/10.1182/blood-2004-08-3280.

[150] Weili Sun and James R Downing. "Haploinsufficiency of AML1 results in a decrease in the number of LTR-HSCs while simultaneously inducing an increase in more mature progenitors". In: *Blood* 104 (12 2004), pp. 3565–3572. URL: https://doi.org/10.1182/blood-2003-12-4349.

[151] Amy C Fan *et al.* "RUNX1 loss renders hematopoietic and leukemic cells dependent on IL-3 and sensitive to JAK inhibition". In: *The Journal of Clinical Investigation* 133 (19 2023). URL: https://doi.org/10.1172/JCI167053.

[152] Xiongwei Cai *et al.* "Runx1 Loss Minimally Impacts Long-Term Hematopoietic Stem Cells". In: *PLOS ONE* 6 (12 2011), e28430. URL: https://doi.org/10.1371/journal.pone.0028430.

[153] Motoshi Ichikawa *et al.* "AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis". In: *Nature Medicine* 10 (3 2004), pp. 299–304. URL: https://doi.org/10.1038/nm997.

[154] K P Ng *et al.* "Runx1 deficiency permits granulocyte lineage commitment but impairs subsequent maturation". In: *Oncogenesis* 2 (11 2013), e78–e78. URL: https://doi.org/10.1038/oncsis.2013.41.

[155] Katia Georgopoulos *et al.* "The ikaros gene is required for the development of all lymphoid lineages". In: *Cell* 79 (1 1994), pp. 143–156. URL: https://doi.org/10.1016/0092-8674(94)90407-3.

[156] Barbara L Kee and Cornelis Murre. "Induction of Early B Cell Factor (EBF) and Multiple B Lineage Genes by the Basic Helix-Loop-Helix Transcription Factor E12". In: *Journal of Experimental Medicine* 188 (4 1998), pp. 699–713. URL: https://doi.org/10.1084/jem.188.4.699.

[157] Haishan Lin and Rudolf Grosschedl. "Failure of B-cell differentiation in mice lacking the transcription factor EBF". In: *Nature* 376 (6537 1995), pp. 263–267. URL: https://doi.org/10.1038/376263a0.

[158] Stephen L Nutt *et al.* "Commitment to the B-lymphoid lineage depends on the transcription factor Pax5". In: *Nature* 401 (6753 1999), pp. 556–562. URL: https://doi.org/10.1038/44076.

[159] Sebastian Carotta *et al.* "The Transcription Factor PU.1 Controls Dendritic Cell Development and Flt3 Cytokine Receptor Expression in a Dose-Dependent Manner". In: *Immunity* 32 (5 2010), pp. 628–641. URL: https://doi.org/10.1016/j.immuni.2010.05.005.

[160] Ansuman T Satpathy *et al.* "Runx1 and Cbfβ regulate the development of Flt3+ dendritic cell progenitors and restrict myeloproliferative disorder". In: *Blood* 123 (19 2014), pp. 2968–2977. URL: https://doi.org/10.1182/blood-2013-11-539643.

[161] Babacar Cisse *et al.* "Transcription Factor E2-2 Is an Essential and Specific Regulator of Plasmacytoid Dendritic Cell Development". In: *Cell* 135 (1 2008), pp. 37–48. URL: https://doi.org/10.1016/j.cell.2008.09.016.

[162] Gregory C Ippolito *et al.* "Dendritic cell fate is determined by BCL11A". In: *Proceedings of the National Academy of Sciences* 111 (11 2014), E998–E1006. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1319228111.

[163] Tomohiko Tamura *et al.* "IFN Regulatory Factor-4 and -8 Govern Dendritic Cell Subset Development and Their Functional Diversity". In: *The Journal of Immunology* 174 (5 2005), pp. 2573–2581. URL: https://doi.org/10.4049/jimmunol.174.5.2573.

[164] Izumi Sasaki *et al.* "Spi-B is critical for plasmacytoid dendritic cell function and development". In: *Blood* 120 (24 2012), pp. 4733–4743. URL: https://doi.org/10.1182/blood-2012-06-436527.

[165] Catherine M Sawai *et al.* "Transcription factor Runx2 controls the development and migration of plasmacytoid dendritic cells". In: *Journal of Experimental Medicine* 210 (11 2013), pp. 2151–2159. URL: https://doi.org/10.1084/jem.20130443.

[166] Hartmut Döhner, Daniel J Weisdorf, and Clara D Bloomfield. "Acute Myeloid Leukemia". In: *New England Journal of Medicine* 373 (12 2015), pp. 1136–1152. URL: https://www.nejm.org/doi/full/10.1056/NEJMra1406184.

[167] Hagop Kantarjian *et al.* "Acute myeloid leukemia: current progress and future directions". In: *Blood Cancer Journal* 11 (2 2021), p. 41. URL: https://doi.org/10.1038/s41408-021-00425-3.

[168] Courtney D DiNardo *et al.* "Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia". In: *New England Journal of Medicine* 383 (7 2020), pp. 617–629. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa2012971.

[169] Keith W Pratz *et al.* "Long-term follow-up of VIALE-A: Venetoclax and azacitidine in chemotherapy-ineligible untreated acute myeloid leukemia". In: *American Journal of Hematology* 99 (4 2024), pp. 615–624. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajh.27246.

[170] The Cancer Genome Atlas Research Network. "Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia". In: *New England Journal of Medicine* 368 (22 2013), pp. 2059–2074. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1301689.

[171] Elli Papaemmanuil *et al.* "Genomic Classification and Prognosis in Acute Myeloid Leukemia". In: *New England Journal of Medicine* 374 (23 2016), pp. 2209–2221. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1516192.

[172] Hartmut Döhner *et al.* "Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN". In: *Blood* 140 (12 2022), pp. 1345–1377. URL: https://doi.org/10.1182/blood.2022016867.

[173] Iñigo Martincorena and Peter J Campbell. "Somatic mutation in cancer and normal cells". In: *Science* 349 (6255 2015), pp. 1483–1489. URL: https://www.science.org/doi/abs/10.1126/science.aab4082.

[174] Sven Turkalj, Felix A Radtke, and Paresh Vyas. "An Overview of Targeted Therapies in Acute Myeloid Leukemia". In: *HemaSphere* 7 (6 2023), e914. URL: https://journals.lww.com/hemasphere/fulltext/2023/06000/an_overview_of_targeted_therapies_in_acute_myeloid.11.aspx.

[175] Natalija Buza-Vidas *et al.* "FLT3 expression initiates in fully multipotent mouse hematopoietic progenitor cells". In: *Blood* 118 (6 2011), pp. 1544–1548. URL: https://doi.org/10.1182/blood-2010-10-316232.

[176] Naval Daver *et al.* "Targeting FLT3 mutations in AML: review of current knowledge and evidence". In: *Leukemia* 33 (2 2019), pp. 299–312. URL: https://doi.org/10.1038/s41375-018-0357-9.

[177] Fumihiko Hayakawa *et al.* "Tandem-duplicated Flt3 constitutively activates STAT5 and MAP kinase and introduces autonomous cell growth in IL-3-dependent cell lines". In: *Oncogene* 19 (5 2000), pp. 624–631. URL: https://doi.org/10.1038/sj.onc.1203354.

[178] Chunaram Choudhary *et al.* "AML-associated Flt3 kinase domain mutations show signal transduction differences compared with Flt3 ITD mutations". In: *Blood* 106 (1 2005), pp. 265–273. URL: https://doi.org/10.1182/blood-2004-07-2942.

[179] Richard M Stone *et al.* "Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation". In: *New England Journal of Medicine* 377 (5 2017), pp. 454–464. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1614359.

[180] Alexander E Perl *et al.* "Gilteritinib or Chemotherapy for Relapsed or Refractory FLT3-Mutated AML". In: *New England Journal of Medicine* 381 (18 2019), pp. 1728–1740. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1902688.

[181] Harry P Erba *et al.* "Quizartinib plus chemotherapy in newly diagnosed patients with FLT3-internal-tandem-duplication-positive acute myeloid leukaemia (QuANTUM-First): a randomised, double-blind, placebo-controlled, phase 3 trial". In: *The Lancet* 401 (10388 2023), pp. 1571–1583. URL: https://doi.org/10.1016/S0140-6736(23)00464-6.

[182] Elaine R Mardis *et al.* "Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome". In: *New England Journal of Medicine* 361 (11 2009), pp. 1058–1066. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa0903840.

[183] Peter Paschka *et al.* "IDH1 and IDH2 Mutations Are Frequent Genetic Alterations in Acute Myeloid Leukemia and Confer Adverse Prognosis in Cytogenetically Normal Acute Myeloid Leukemia With NPM1 Mutation Without FLT3 Internal Tandem Duplication". In: *Journal of Clinical Oncology* 28 (22 2010), pp. 3636–3643. URL: https://ascopubs.org/doi/abs/10.1200/JCO.2010.28.3762.

[184] Lenny Dang *et al.* "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate". In: *Nature* 462 (7274 2009), pp. 739–744. URL: https://doi.org/10.1038/nature08617.

[185] Patrick S Ward *et al.* "The Common Feature of Leukemia-Associated IDH1 and IDH2 Mutations Is a Neomorphic Enzyme Activity Converting α-Ketoglutarate to 2-Hydroxyglutarate". In: *Cancer Cell* 17 (3 2010), pp. 225–234. URL: https://www.sciencedirect.com/science/article/pii/S153561081000036X.

[186] Wei Xu *et al.* "Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases". In: *Cancer Cell* 19 (1 2011), pp. 17–30. URL: https://doi.org/10.1016/j.ccr.2010.12.014.

[187] Chao Lu *et al.* "IDH mutation impairs histone demethylation and results in a block to cell differentiation". In: *Nature* 483 (7390 2012), pp. 474–478. URL: https://doi.org/10.1038/nature10860.

[188] Maria E Figueroa *et al.* "Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation". In: *Cancer Cell* 18 (6 2010), pp. 553–567. URL: https://doi.org/10.1016/j.ccr.2010.11.015.

[189] Courtney D DiNardo *et al.* "Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML". In: *New England Journal of Medicine* 378 (25 2018), pp. 2386–2398. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1716984.

[190] Eytan M Stein *et al.* "Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia". In: *Blood* 130 (6 2017), pp. 722–731. URL: https://doi.org/10.1182/blood-2017-04-779405.

[191] Stéphane de Botton *et al.* "Enasidenib vs conventional care in older patients with late-stage mutant-IDH2 relapsed/refractory AML: a randomized phase 3 trial". In: *Blood* 141 (2 2023), pp. 156–167. URL: https://doi.org/10.1182/blood.2021014901.

[192] Pau Montesinos *et al.* "Ivosidenib and Azacitidine in IDH1-Mutated Acute Myeloid Leukemia". In: *New England Journal of Medicine* 386 (16 2022), pp. 1519–1531. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa2117344.

[193] Courtney D DiNardo *et al.* "Enasidenib plus azacitidine versus azacitidine alone in patients with newly diagnosed, mutant-IDH2 acute myeloid leukaemia (AG221-AML-005): a single-arm, phase 1b and randomised, phase 2 trial". In: *The Lancet Oncology* 22 (11 2021), pp. 1597–1608. URL: https://doi.org/10.1016/S1470-2045(21)00494-0.

[194] Marina Konopleva *et al.* "Mechanisms of apoptosis sensitivity and resistance to the BH3 mimetic ABT-737 in acute myeloid leukemia". In: *Cancer Cell* 10 (5 2006), pp. 375–388. URL: https://www.sciencedirect.com/science/article/pii/S1535610806003138.

[195] Eleni D Lagadinou *et al.* "BCL-2 Inhibition Targets Oxidative Phosphorylation and Selectively Eradicates Quiescent Human Leukemia Stem Cells". In: *Cell Stem Cell* 12 (3 2013), pp. 329–341. URL: https://www.sciencedirect.com/science/article/pii/S1934590912007552.

[196] Marina Konopleva *et al.* "Efficacy and Biological Correlates of Response in a Phase II Study of Venetoclax Monotherapy in Patients with Acute Myelogenous Leukemia". In: *Cancer Discovery* 6 (10 2016), pp. 1106–1117. URL: https://doi.org/10.1158/2159-8290.CD-16-0313.

[197] Daniel A Pollyea *et al.* "Impact of Venetoclax and Azacitidine in Treatment-Naïve Patients with Acute Myeloid Leukemia and IDH1/2 Mutations". In: *Clinical Cancer Research* 28 (13 2022), pp. 2753–2761. URL: https://doi.org/10.1158/1078-0432.CCR-21-3467.

[198] Steven M Chan *et al.* "Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia". In: *Nature Medicine* 21 (2 2015), pp. 178–184. URL: https://doi.org/10.1038/nm.3788.

[199] Curtis A Lachowiez *et al.* "A Phase Ib/II Study of Ivosidenib with Venetoclax ± Azacitidine in IDH1-Mutated Myeloid Malignancies". In: *Blood Cancer Discovery* 4 (4 2023), pp. 276–293. URL: https://doi.org/10.1158/2643-3230.BCD-22-0205.

[200] Thomas A Milne *et al.* "MLL associates specifically with a subset of transcriptionally active target genes". In: *Proceedings of the National Academy of Sciences* 102 (41 2005), pp. 14765–14770. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0503630102.

[201] Andrei V Krivtsov *et al.* "Transformation from committed progenitor to leukaemia stem cell initiated by MLL–AF9". In: *Nature* 442 (7104 2006), pp. 818–822. URL: https://doi.org/10.1038/nature04980.

[202] Akihiko Yokoyama *et al.* "The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis". In: *Cell* 123 (2 2005), pp. 207–218. URL: https://www.sciencedirect.com/science/article/pii/S0092867405009797.

[203] Thomas A Milne *et al.* "Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors". In: *Proceedings of the National Academy of Sciences* 102 (3 2005), pp. 749–754. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0408836102.

[204] Christina M Hughes *et al.* "Menin Associates with a Trithorax Family Histone Methyltransferase Complex and with the Hoxc8 Locus". In: *Molecular Cell* 13 (4 2004), pp. 587–597. URL: https://www.sciencedirect.com/science/article/pii/S1097276504000814.

[205] Yadira M Soto-Feliciano *et al.* "A Molecular Switch between Mammalian MLL Complexes Dictates Response to Menin–MLL Inhibition". In: *Cancer Discovery* 13 (1 2023), pp. 146–169. URL: https://doi.org/10.1158/2159-8290.CD-22-0416.

[206] Jolanta Grembecka *et al.* "Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia". In: *Nature Chemical Biology* 8 (3 2012), pp. 277–284. URL: https://doi.org/10.1038/nchembio.773.

[207] Dmitry Borkin *et al.* "Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo". In: *Cancer Cell* 27 (4 2015), pp. 589–602. URL: https://www.sciencedirect.com/science/article/pii/S1535610815000665.

[208] Andrei V Krivtsov *et al.* "A Menin-MLL Inhibitor Induces Specific Chromatin Changes and Eradicates Disease in Models of MLL-Rearranged Leukemia". In: *Cancer Cell* 36 (6 2019), 660–673.e11. URL: https://www.sciencedirect.com/science/article/pii/S1535610819305215.

[209] Hannah J Uckelmann *et al.* "Therapeutic targeting of preleukemia cells in a mouse model of NPM1 mutant acute myeloid leukemia". In: *Science* 367 (6477 2020), pp. 586–590. URL: https://www.science.org/doi/abs/10.1126/science.aax5863.

[210] Michael W M Kühn *et al.* "Targeting Chromatin Regulators Inhibits Leukemogenic Gene Expression in NPM1 Mutant Leukemia". In: *Cancer Discovery* 6 (10 2016), pp. 1166–1181. URL: https://doi.org/10.1158/2159-8290.CD-16-0237.

[211] Hannah J Uckelmann *et al.* "Mutant NPM1 Directly Regulates Oncogenic Transcription in Acute Myeloid Leukemia". In: *Cancer Discovery* 13 (3 2023), pp. 746–765. URL: https://doi.org/10.1158/2159-8290.CD-22-0366.

[212] Xue Qing David Wang *et al.* "Mutant NPM1 Hijacks Transcriptional Hubs to Maintain Pathogenic Gene Programs in Acute Myeloid Leukemia". In: *Cancer Discovery* 13 (3 2023), pp. 724–745. URL: https://doi.org/10.1158/2159-8290.CD-22-0424.

[213] Ghayas C Issa *et al.* "The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia". In: *Nature* 615 (7954 2023), pp. 920–924. URL: https://doi.org/10.1038/s41586-023-05812-3.

[214] Ghayas C Issa *et al.* "Early Results of the Phase I/II Study Investigating the All-Oral Combination of the Menin Inhibitor Revumenib (SNDX-5613) with Decitabine/Cedazuridine (ASTX727) and Venetoclax in Acute Myeloid Leukemia (SAVE)". In: *Blood* 142 (Supplement 1 2023), p. 58. URL: https://doi.org/10.1182/blood-2023-182337.

[215] Elias Jabbour *et al.* "A First-in-Human Phase 1 Study of the Menin-KMT2A (MLL1) Inhibitor JNJ-75276617 in Adult Patients with Relapsed/Refractory Acute Leukemia Harboring KMT2A or NPM1 Alterations". In: *Blood* 142 (Supplement 1 2023), p. 57. URL: https://doi.org/10.1182/blood-2023-172422.

[216] Harry Erba *et al.* "AML-475 Activity, Tolerability, and Resistance Profile of the Menin Inhibitor Ziftomenib in Adults With Relapsed/Refractory NPM1-Mutated AML". In: *Clinical Lymphoma Myeloma and Leukemia* 23 (2023), S304–S305. URL: https://www.sciencedirect.com/science/article/pii/S2152265023010674.

[217] Bayard Clarkson *et al.* "Studies of cellular proliferation in human leukemia. III. Behavior of leukemic cells in three adults with acute leukemia given continuous infusions of H-thymidine for 8 or 10 days". In: *Cancer* 25 (6 1970), pp. 1237–1260. URL: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%28197006%2925%3A6%3C1237%3A%3AAID-CNCR2820250602%3E3.0.CO%3B2-7.

[218] James D Griffin and Bob Löwenberg. "Clonogenic Cells in Acute Myeloblastic Leukemia". In: *Blood* 68 (6 1986), pp. 1185–1195. URL: https://www.sciencedirect.com/science/article/pii/S0006497120859382.

[219] John E Dick. "Stem cell concepts renew cancer research". In: *Blood* 112 (13 2008), pp. 4793–4807. URL: https://doi.org/10.1182/blood-2008-08-077941.

[220] K D Sabbath *et al.* "Heterogeneity of clonogenic cells in acute myeloblastic leukemia". In: *The Journal of Clinical Investigation* 75 (2 1985), pp. 746–753. URL: https://doi.org/10.1172/JCI111756.

[221] R N Buick, M D Minden, and E A McCulloch. "Self-Renewal in Culture of Proliferative Blast Progenitor Cells in Acute Myeloblasts Leukemia". In: *Blood* 54 (1 1979), pp. 95–104. URL: https://www.sciencedirect.com/science/article/pii/S0006497120726327.

[222] Tsvee Lapidot *et al.* "A cell initiating human acute myeloid leukaemia after transplantation into SCID mice". In: *Nature* 367 (6464 1994), pp. 645–648. URL: https://doi.org/10.1038/367645a0.

[223] Dominique Bonnet and John E Dick. "Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell". In: *Nature Medicine* 3 (7 1997), pp. 730–737. URL: https://doi.org/10.1038/nm0797-730.

[224] Fumihiko Ishikawa *et al.* "Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region". In: *Nature Biotechnology* 25 (11 2007), pp. 1315–1321. URL: https://doi.org/10.1038/nbt1350.

[225] Yinghui Guan, Brigitte Gerhard, and Donna E Hogge. "Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML)". In: *Blood* 101 (8 2003), pp. 3142–3149. URL: https://doi.org/10.1182/blood-2002-10-3062.

[226] Kristin J Hope, Liqing Jin, and John E Dick. "Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity". In: *Nature Immunology* 5 (7 2004), pp. 738–743. URL: https://doi.org/10.1038/ni1080.

[227] A Blair *et al.* "Lack of Expression of Thy-1 (CD90) on Acute Myeloid Leukemia Cells With Long-Term Proliferative Ability In Vitro and In Vivo". In: *Blood* 89 (9 1997), pp. 3104–3112. URL: https://www.sciencedirect.com/science/article/pii/S0006497120583201.

[228] David C Taussig *et al.* "Anti-CD38 antibody–mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells". In: *Blood* 112 (3 2008), pp. 568–575. URL: https://doi.org/10.1182/blood-2007-10-118331.

[229] A Blair, D E Hogge, and H J Sutherland. "Most Acute Myeloid Leukemia Progenitor Cells With Long-Term Proliferative Ability In Vitro and In Vivo Have the Phenotype CD34+/CD71-/HLA-DR-". In: *Blood* 92 (11 1998), pp. 4325–4335. URL: https://www.sciencedirect.com/science/article/pii/S0006497120577665.

[230] David C Taussig *et al.* "Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34- fraction". In: *Blood* 115 (10 2010), pp. 1976–1984. URL: https://doi.org/10.1182/blood-2009-02-206565.

[231] Kolja Eppert *et al.* "Stem cell gene expression programs influence clinical outcome in human leukemia". In: *Nature Medicine* 17 (9 2011), pp. 1086–1093. URL: https://doi.org/10.1038/nm.2415.

[232] Jean-Emmanuel Sarry *et al.* "Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice". In: *The Journal of Clinical Investigation* 121 (1 2011), pp. 384–395. URL: https://doi.org/10.1172/JCI41495.

[233] Tzu-Chieh Ho *et al.* "Evolution of acute myelogenous leukemia stem cell properties after treatment and progression". In: *Blood* 128 (13 2016), pp. 1671–1678. URL: https://doi.org/10.1182/blood-2016-02-695312.

[234] Shanshan Pei *et al.* "A Novel Type of Monocytic Leukemia Stem Cell Revealed by the Clinical Use of Venetoclax-Based Therapy". In: *Cancer Discovery* 13 (9 2023), pp. 2032–2049. URL: https://doi.org/10.1158/2159-8290.CD-22-1297.

[235] Namyoung Jung *et al.* "An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis". In: *Nature Communications* 6 (1 2015), p. 8489. URL: https://doi.org/10.1038/ncomms9489.

[236] Stanley W K Ng *et al.* "A 17-gene stemness score for rapid determination of risk in acute leukaemia". In: *Nature* 540 (7633 2016), pp. 433–437. URL: https://doi.org/10.1038/nature20598.

[237] Antonio Cozzio *et al.* "Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors". In: *Genes Development* 17 (24 2003), pp. 3029–3035. URL: http://genesdev.cshlp.org/content/17/24/3029.abstract.

[238] Brian J P Huntly *et al.* "MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors". In: *Cancer Cell* 6 (6 2004), pp. 587–596. URL: https://www.sciencedirect.com/science/article/pii/S1535610804003101.

[239] Chi Wai So *et al.* "MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice". In: *Cancer Cell* 3 (2 2003), pp. 161–171. URL: https://www.sciencedirect.com/science/article/pii/S1535610803000199.

[240] Ross Kinstrie *et al.* "Heterogeneous leukemia stem cells in myeloid blast phase chronic myeloid leukemia". In: *Blood Advances* 1 (3 2016), pp. 160–169. URL: https://doi.org/10.1182/bloodadvances.2016000810.

[241] Wendy W Pang *et al.* "Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes". In: *Proceedings of the National Academy of Sciences* 110 (8 2013), pp. 3011–3016. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1222861110.

[242] Britta Will *et al.* "Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations". In: *Blood* 120 (10 2012), pp. 2076–2086. URL: https://doi.org/10.1182/blood-2011-12-399683.

[243] Petter S Woll *et al.* "Myelodysplastic Syndromes Are Propagated by Rare and Distinct Human Cancer Stem Cells In Vivo". In: *Cancer Cell* 25 (6 2014), pp. 794–808. URL: https://www.sciencedirect.com/science/article/pii/S1535610814001342.

[244] Andrew J Gentles *et al.* "Association of a Leukemic Stem Cell Gene Expression Signature With Clinical Outcomes in Acute Myeloid Leukemia". In: *JAMA* 304 (24 2010), pp. 2706–2715. URL: https://doi.org/10.1001/jama.2010.1862.

[245] Shanshan Pei *et al.* "AMPK/FIS1-Mediated Mitophagy Is Required for Self-Renewal of Human AML Stem Cells". In: *Cell Stem Cell* 23 (1 2018), 86–100.e6. URL: https://www.sciencedirect.com/science/article/pii/S193459091830239X.

[246] Jenny Yeung *et al.* "β-Catenin Mediates the Establishment and Drug Resistance of MLL Leukemic Stem Cells". In: *Cancer Cell* 18 (6 2010), pp. 606–618. URL: https://www.sciencedirect.com/science/article/pii/S1535610810004344.

[247] Yingzi Wang *et al.* "The Wnt/β-Catenin Pathway Is Required for the Development of Leukemia Stem Cells in AML". In: *Science* 327 (5973 2010), pp. 1650–1653. URL: https://www.science.org/doi/abs/10.1126/science.1186624.

[248] Chen Zhao *et al.* "Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia". In: *Nature* 458 (7239 2009), pp. 776–779. URL: https://doi.org/10.1038/nature07737.

[249] Anil Sadarangani *et al.* "GLI2 inhibition abrogates human leukemia stem cell dormancy". In: *Journal of Translational Medicine* 13 (1 2015), p. 98. URL: https://doi.org/10.1186/s12967-015-0453-9.

[250] Monica L Guzman *et al.* "Nuclear factor-κB is constitutively activated in primitive human acute myelogenous leukemia cells". In: *Blood* 98 (8 2001), pp. 2301–2307. URL: https://doi.org/10.1182/blood.V98.8.2301.

[251] Peter van Galen *et al.* "Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity". In: *Cell* 176 (6 2019), 1265–1281.e24. URL: https://www.sciencedirect.com/science/article/pii/S0092867419300947.

[252] Sergi Beneyto-Calabuig *et al.* "Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia". In: *Cell Stem Cell* 30 (5 2023), 706–721.e8. URL: https://www.sciencedirect.com/science/article/pii/S1934590923001194.

[253] Lars Velten *et al.* "Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics". In: *Nature Communications* 12 (1 2021), p. 1366. URL: https://doi.org/10.1038/s41467-021-21650-1.

[254] Andy G X Zeng *et al.* "A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia". In: *Nature Medicine* 28 (6 2022), pp. 1212–1223. URL: https://doi.org/10.1038/s41591-022-01819-x.

[255] Wim Terpstra *et al.* "Fluorouracil Selectively Spares Acute Myeloid Leukemia Cells With Long-Term Growth Abilities in Immunodeficient Mice and in Culture". In: *Blood* 88 (6 1996), pp. 1944–1950. URL: https://www.sciencedirect.com/science/article/pii/S0006497120623916.

[256] Matteo Maria Naldini *et al.* "Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia". In: *Nature Communications* 14 (1 2023), p. 1285. URL: https://doi.org/10.1038/s41467-023-36969-0.

[257] C Craddock *et al.* "Azacitidine fails to eradicate leukemic stem/progenitor cell populations in patients with acute myeloid leukemia and myelodysplasia". In: *Leukemia* 27 (5 2013), pp. 1028–1036. URL: https://doi.org/10.1038/leu.2012.312.

[258] Sumiko Takao *et al.* "Epigenetic mechanisms controlling human leukemia stem cells and therapy resistance". In: *bioRxiv* (2024), p. 2022.09.22.509005. URL: https://www.biorxiv.org/content/biorxiv/early/2024/07/26/2022.09.22.509005.full.pdf.

[259] Charles F Craddock *et al.* "Outcome of Azacitidine Therapy in Acute Myeloid Leukemia Is not Improved by Concurrent Vorinostat Therapy but Is Predicted by a Diagnostic Molecular Signature". In: *Clinical Cancer Research* 23 (21 2017), pp. 6430–6440. URL: https://doi.org/10.1158/1078-0432.CCR-17-1423.

[260] Nicholas J Short *et al.* "Association of Measurable Residual Disease With Survival Outcomes in Patients With Acute Myeloid Leukemia: A Systematic Review and Meta-analysis". In: *JAMA Oncology* 6 (12 2020), pp. 1890–1899. URL: https://doi.org/10.1001/jamaoncol.2020.4600.

[261] Keith W Pratz *et al.* "Measurable Residual Disease Response and Prognosis in Treatment-Naïve Acute Myeloid Leukemia With Venetoclax and Azacitidine". In: *Journal of Clinical Oncology* 40 (8 2022), pp. 855–865. URL: https://ascopubs.org/doi/abs/10.1200/JCO.21.01546.

[262] Mojca Jongen-Lavrencic *et al.* "Molecular Minimal Residual Disease in Acute Myeloid Leukemia". In: *New England Journal of Medicine* 378 (13 2018), pp. 1189–1199. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1716863.

[263] Jan Krönke *et al.* "Monitoring of Minimal Residual Disease in NPM1-Mutated Acute Myeloid Leukemia: A Study From the German-Austrian Acute Myeloid Leukemia Study Group". In: *Journal of Clinical Oncology* 29 (19 2011), pp. 2709–2716. URL: https://ascopubs.org/doi/abs/10.1200/JCO.2011.35.0371.

[264] David Vetrie, G Vignir Helgason, and Mhairi Copland. "The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML". In: *Nature Reviews Cancer* 20 (3 2020), pp. 158–173. URL: https://doi.org/10.1038/s41568-019-0230-9.

[265] Yoshikane Kikushige *et al.* "TIM-3 Is a Promising Target to Selectively Kill Acute Myeloid Leukemia Stem Cells". In: *Cell Stem Cell* 7 (6 2010), pp. 708–717. URL: https://www.sciencedirect.com/science/article/pii/S1934590910006363.

[266] Yoshikane Kikushige *et al.* "A TIM-3/Gal-9 Autocrine Stimulatory Loop Drives Self-Renewal of Human Myeloid Leukemia Stem Cells and Leukemic Progression". In: *Cell Stem Cell* 17 (3 2015), pp. 341–352. URL: https://www.sciencedirect.com/science/article/pii/S1934590915003100.

[267] Max Jan *et al.* "Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker". In: *Proceedings of the National Academy of Sciences* 108 (12 2011), pp. 5009–5014. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1100551108.

[268] Siddhartha Jaiswal *et al.* "CD47 Is Upregulated on Circulating Hematopoietic Stem Cells and Leukemia Cells to Avoid Phagocytosis". In: *Cell* 138 (2 2009), pp. 271–285. URL: https://www.sciencedirect.com/science/article/pii/S0092867409006515.

[269] Naoki Hosen *et al.* "CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia". In: *Proceedings of the National Academy of Sciences* 104 (26 2007), pp. 11008–11013. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0704271104.

[270] Lynn Quek *et al.* "Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib". In: *Nature Medicine* 24 (8 2018), pp. 1167–1177. URL: https://doi.org/10.1038/s41591-018-0115-6.

[271] John S Welch *et al.* "The Origin and Evolution of Mutations in Acute Myeloid Leukemia". In: *Cell* 150 (2 2012), pp. 264–278. URL: https://www.sciencedirect.com/science/article/pii/S0092867412007775.

[272] Li Ding *et al.* "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing". In: *Nature* 481 (7382 2012), pp. 506–510. URL: https://doi.org/10.1038/nature10738.

[273] Pierre Hirsch *et al.* "Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukaemia". In: *Nature Communications* 7 (1 2016), p. 12475. URL: https://doi.org/10.1038/ncomms12475.

[274] Kiyomi Morita *et al.* "Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics". In: *Nature Communications* 11 (1 2020), p. 5327. URL: https://doi.org/10.1038/s41467-020-19119-8.

[275] Linde A Miles *et al.* "Single-cell mutation analysis of clonal evolution in myeloid malignancies". In: *Nature* 587 (7834 2020), pp. 477–482. URL: https://doi.org/10.1038/s41586-020-2864-x.

[276] Jeffery M Klco *et al.* "Functional Heterogeneity of Genetically Defined Subclones in Acute Myeloid Leukemia". In: *Cancer Cell* 25 (3 2014), pp. 379–392. URL: https://www.sciencedirect.com/science/article/pii/S1535610814000543.

[277] Toshihiro Miyamoto, Irving L Weissman, and Koichi Akashi. "AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation". In: *Proceedings of the National Academy of Sciences* 97 (13 2000), pp. 7521–7526. URL: https://www.pnas.org/doi/abs/10.1073/pnas.97.13.7521.

[278] Max Jan *et al.* "Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia". In: *Science Translational Medicine* 4 (149 2012), 149ra118–149ra118. URL: https://www.science.org/doi/abs/10.1126/scitranslmed.3004315.

[279] M Ryan Corces-Zimmerman *et al.* "Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission". In: *Proceedings of the National Academy of Sciences* 111 (7 2014), pp. 2548–2553. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1324297111.

[280] Liran I Shlush *et al.* "Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia". In: *Nature* 506 (7488 2014), pp. 328–333. URL: https://doi.org/10.1038/nature13038.

[281] Siddhartha Jaiswal *et al.* "Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes". In: *New England Journal of Medicine* 371 (26 2014), pp. 2488–2498. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1408617.

[282] Giulio Genovese *et al.* "Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence". In: *New England Journal of Medicine* 371 (26 2014), pp. 2477–2487. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1409405.

[283] Mingchao Xie *et al.* "Age-related mutations associated with clonal hematopoietic expansion and malignancies". In: *Nature Medicine* 20 (12 2014), pp. 1472–1478. URL: https://doi.org/10.1038/nm.3733.

[284] Hui Yang *et al.* "Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies". In: *Blood* 131 (3 2018), pp. 328–341. URL: https://doi.org/10.1182/blood-2017-06-789669.

[285] Thomas Köhnke *et al.* "Human ASXL1-Mutant Hematopoiesis Is Driven by a Truncated Protein Associated with Aberrant Deubiquitination of H2AK119". In: *Blood Cancer Discovery* 5 (3 2024), pp. 202–223. URL: https://doi.org/10.1158/2643-3230.BCD-23-0235.

[286] Omar Abdel-Wahab *et al.* "ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression". In: *Cancer Cell* 22 (2 2012), pp. 180–193. URL: https://www.sciencedirect.com/science/article/pii/S1535610812003030.

[287] Shuhei Asada *et al.* "Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis". In: *Nature Communications* 9 (1 2018), p. 2733. URL: https://doi.org/10.1038/s41467-018-05085-9.

[288] Daniel Hormaechea-Agulla *et al.* "Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling". In: *Cell Stem Cell* 28 (8 2021), 1428–1442.e6. URL: https://www.sciencedirect.com/science/article/pii/S1934590921001089.

[289] Marlies Meisel *et al.* "Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host". In: *Nature* 557 (7706 2018), pp. 580–584. URL: https://doi.org/10.1038/s41586-018-0125-z.

[290] Zhigang Cai *et al.* "Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis". In: *Cell Stem Cell* 23 (6 2018), 833–849.e5. URL: https://www.sciencedirect.com/science/article/pii/S1934590918304946.

[291] Jennifer M SanMiguel *et al.* "Distinct Tumor Necrosis Factor Alpha Receptors Dictate Stem Cell Fitness versus Lineage Output in Dnmt3a-Mutant Clonal Hematopoiesis". In: *Cancer Discovery* 12 (12 2022), pp. 2763–2773. URL: https://doi.org/10.1158/2159-8290.CD-22-0086.

[292] Francisco Caiado *et al.* "Aging drives Tet2+/- clonal hematopoiesis via IL-1 signaling". In: *Blood* 141 (8 2023), pp. 886–903. URL: https://doi.org/10.1182/blood.2022016835.

[293] Matthew A Loberg *et al.* "Sequentially inducible mouse models reveal that Npm1 mutation causes malignant transformation of Dnmt3a-mutant clonal hematopoiesis". In: *Leukemia* 33 (7 2019), pp. 1635–1649. URL: https://doi.org/10.1038/s41375-018-0368-6.

[294] Robert L Bowman *et al.* "Modeling clonal evolution and oncogenic dependency in vivo in the context of hematopoietic transformation". In: *bioRxiv* (2022), p. 2022.05.18.492524. URL: https://www.biorxiv.org/content/biorxiv/early/2022/05/18/2022.05.18.492524.full.pdf.

[295] Alan H Shih *et al.* "Mutational Cooperativity Linked to Combinatorial Epigenetic Gain of Function in Acute Myeloid Leukemia". In: *Cancer Cell* 27 (4 2015), pp. 502–515. URL: https://www.sciencedirect.com/science/article/pii/S1535610815000975.

[296] Vera Grossmann *et al.* "Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype". In: *Blood* 118 (23 2011), pp. 6153–6163. URL: https://doi.org/10.1182/blood-2011-07-365320.

[297] V I Gaidzik *et al.* "RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features". In: *Leukemia* 30 (11 2016). (AMLSG), pp. 2160–2168. URL: https://doi.org/10.1038/leu.2016.126.

[298] Anna L Brown, Christopher N Hahn, and Hamish S Scott. "Secondary leukemia in patients with germline transcription factor mutations (RUNX1, GATA2, CEBPA)". In: *Blood* 136 (1 2020), pp. 24–35. URL: https://doi.org/10.1182/blood.2019000937.

[299] Anna L Brown *et al.* "RUNX1-mutated families show phenotype heterogeneity and a somatic mutation profile unique to germline predisposed AML". In: *Blood Advances* 4 (6 2020), pp. 1131–1144. URL: https://doi.org/10.1182/bloodadvances.2019000901.

[300] Florian Heidel *et al.* "Clinical resistance to the kinase inhibitor PKC412 in acute myeloid leukemia by mutation of Asn-676 in the FLT3 tyrosine kinase domain". In: *Blood* 107 (1 2006), pp. 293–300. URL: https://doi.org/10.1182/blood-2005-06-2469.

[301] Obdulio Piloto *et al.* "Prolonged exposure to FLT3 inhibitors leads to resistance via activation of parallel signaling pathways". In: *Blood* 109 (4 2006), pp. 1643–1652. URL: https://doi.org/10.1182/blood-2006-05-023804.

[302] Feng Wang *et al.* "Leukemia stemness and co-occurring mutations drive resistance to IDH inhibitors in acute myeloid leukemia". In: *Nature Communications* 12 (1 2021), p. 2607. URL: https://doi.org/10.1038/s41467-021-22874-x.

[303] Andrew M Intlekofer *et al.* "Acquired resistance to IDH inhibition through trans or cis dimer-interface mutations". In: *Nature* 559 (7712 2018), pp. 125–129. URL: https://doi.org/10.1038/s41586-018-0251-7.

[304] James J Harding *et al.* "Isoform Switching as a Mechanism of Acquired Resistance to Mutant Isocitrate Dehydrogenase Inhibition". In: *Cancer Discovery* 8 (12 2018), pp. 1540–1547. URL: https://doi.org/10.1158/2159-8290.CD-18-0877.

[305] Brenda Chyla *et al.* "Genetic biomarkers of sensitivity and resistance to venetoclax monotherapy in patients with relapsed acute myeloid leukemia". In: *American Journal of Hematology* 93 (8 2018), E202–E205. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajh.25146.

[306] Goichi Yoshimoto *et al.* "FLT3-ITD up-regulates MCL-1 to promote survival of stem cells in acute myeloid leukemia via FLT3-ITD–specific STAT5 activation". In: *Blood* 114 (24 2009), pp. 5034–5043. URL: https://doi.org/10.1182/blood-2008-12-196055.

[307] Ong Faustine, Kim Kunhwa, and Y Konopleva Marina. "Venetoclax resistance: mechanistic insights and future strategies". In: *Cancer Drug Resistance* 5 (2 2022), pp. 380–400. URL: https://www.oaepublish.com/articles/cdr.2021.125.

[308] Florian Perner *et al.* "MEN1 mutations mediate clinical resistance to menin inhibition". In: *Nature* 615 (7954 2023), pp. 913–919. URL: https://doi.org/10.1038/s41586-023-05755-9.

[309] Brian Parkin *et al.* "Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia". In: *Blood* 121 (2 2013), pp. 369–377. URL: https://doi.org/10.1182/blood-2012-04-427039.

[310] Manoj Garg *et al.* "Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse". In: *Blood* 126 (22 2015), pp. 2491–2501. URL: https://doi.org/10.1182/blood-2015-05-646240.

[311] Maurizio Pellegrino *et al.* "High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics". In: *Genome Research* 28 (9 2018), pp. 1345–1352. URL: http://genome.cshlp.org/content/28/9/1345.abstract.

[312] Allegra A Petti *et al.* "Genetic and Transcriptional Contributions to Relapse in Normal Karyotype Acute Myeloid Leukemia". In: *Blood Cancer Discovery* 3 (1 2022), pp. 32–49. URL: https://doi.org/10.1158/2643-3230.BCD-21-0050.

[313] Liran I Shlush *et al.* "Tracing the origins of relapse in acute myeloid leukaemia to stem cells". In: *Nature* 547 (7661 2017), pp. 104–108. URL: https://doi.org/10.1038/nature22993.

[314] Kevin Nuno *et al.* "Convergent epigenetic evolution drives relapse in acute myeloid leukemia". In: *eLife* 13 (2024). Ed. by Amit K Verma, Tony Ng, and Amit K Verma, e93019. URL: https://doi.org/10.7554/eLife.93019.

[315] Elizabeth Heyes *et al.* "TET2 lesions enhance the aggressiveness of CEBPA-mutant acute myeloid leukemia by rebalancing GATA2 expression". In: *Nature Communications* 14 (1 2023), p. 6185. URL: https://doi.org/10.1038/s41467-023-41927-x.

[316] Liubin Yang *et al.* "DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias". In: *Cancer Cell* 29 (6 2016), pp. 922–934. URL: https://www.sciencedirect.com/science/article/pii/S1535610816302082.

[317] Min Yang *et al.* "A unique role of p53 haploinsufficiency or loss in the development of acute myeloid leukemia with FLT3-ITD mutation". In: *Leukemia* 36 (3 2022), pp. 675–686. URL: https://doi.org/10.1038/s41375-021-01452-6.

[318] Haiyang Yun *et al.* "Mutational synergy during leukemia induction remodels chromatin accessibility, histone modifications and three-dimensional DNA topology to alter gene expression". In: *Nature Genetics* 53 (10 2021), pp. 1443–1455. URL: https://doi.org/10.1038/s41588-021-00925-9.

[319] Kendra K Jackson *et al.* "BCOR Loss Confers Increased Stemness and Partially Rescues RUNX1-Deficient Phenotypes in Human Hematopoietic Stem and Progenitor Cells". In: *Blood* 142 (Supplement 1 2023), p. 946. URL: https://doi.org/10.1182/blood-2023-187865.

[320] A V Krivtsov *et al.* "Cell of origin determines clinically relevant subtypes of MLL-rearranged AML". In: *Leukemia* 27 (4 2013), pp. 852–860. URL: https://doi.org/10.1038/leu.2012.363.

[321] Iain C Macaulay *et al.* "GT-seq: parallel sequencing of single-cell genomes and transcriptomes". In: *Nature Methods* 12 (6 2015), pp. 519–522. URL: https://doi.org/10.1038/nmeth.3370.

[322] Itay Tirosh *et al.* "Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma". In: *Nature* 539 (7628 2016), pp. 309–313. URL: https://doi.org/10.1038/nature20123.

[323] Mariella G Filbin *et al.* "Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq". In: *Science* 360 (6386 2018), pp. 331–335. URL: https://www.science.org/doi/abs/10.1126/science.aao4750.

[324] Alice Giustacchini *et al.* "Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia". In: *Nature Medicine* 23 (6 2017), pp. 692–702. URL: https://doi.org/10.1038/nm.4336.

[325] Anna S Nam *et al.* "Somatic mutations and cell identity linked by Genotyping of Transcriptomes". In: *Nature* 571 (7765 2019), pp. 355–360. URL: https://doi.org/10.1038/s41586-019-1367-0.

[326] Federico Gaiti *et al.* "Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia". In: *Nature* 569 (7757 2019), pp. 576–580. URL: https://doi.org/10.1038/s41586-019-1198-z.

[327] Anna S Nam *et al.* "Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation". In: *Nature Genetics* 54 (10 2022), pp. 1514–1526. URL: https://doi.org/10.1038/s41588-022-01179-9.

[328] Mariela Cortés-López *et al.* "Single-cell multi-omics defines the cell-type-specific impact of splicing aberrations in human hematopoietic clonal outgrowths". In: *Cell Stem Cell* 30 (9 2023), 1262–1281.e8. URL: https://www.sciencedirect.com/science/article/pii/S1934590923002576.

[329] Alba Rodriguez-Meira *et al.* "Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing". In: *Molecular Cell* 73 (6 2019), 1292–1305.e8. URL: https://www.sciencedirect.com/science/article/pii/S1097276519300097.

[330] Alba Rodriguez-Meira *et al.* "TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing". In: *STAR Protocols* 1 (3 2020), p. 100125. URL: https://www.sciencedirect.com/science/article/pii/S266616672030112X.

[331] Alba Rodriguez-Meira *et al.* "Single-cell multi-omics identifies chronic inflammation as a driver of TP53-mutant leukemic evolution". In: *Nature Genetics* 55 (9 2023), pp. 1531–1541. URL: https://doi.org/10.1038/s41588-023-01480-1.

[332] Sven Turkalj *et al.* "GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells". In: *Cell Stem Cell* 30 (5 2023), 722–740.e11. URL: https://www.sciencedirect.com/science/article/pii/S1934590923001303.

[333] Sven Turkalj *et al.* "A protocol for simultaneous high-sensitivity genotyping and chromatin accessibility profiling in single cells". In: *STAR Protocols* 4 (4 2023), p. 102641. URL: https://www.sciencedirect.com/science/article/pii/S2666166723006081.

[334] Heng Li and Richard Durbin. "Fast and accurate short read alignment with Burrows–Wheeler transform". In: *Bioinformatics* 25 (14 2009), pp. 1754–1760. URL: https://doi.org/10.1093/bioinformatics/btp324.

[335] Kristian Cibulskis *et al.* "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples". In: *Nature Biotechnology* 31 (3 2013), pp. 213–219. URL: https://doi.org/10.1038/nbt.2514.

[336] Sangtae Kim *et al.* "Strelka2: fast and accurate calling of germline and somatic variants". In: *Nature Methods* 15 (8 2018), pp. 591–594. URL: https://doi.org/10.1038/s41592-018-0051-x.

[337] Zhongwu Lai *et al.* "VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research". In: *Nucleic Acids Research* 44 (11 2016), e108–e108. URL: https://doi.org/10.1093/nar/gkw227.

[338] Kai Ye *et al.* "Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads". In: *Bioinformatics* 25 (21 2009), pp. 2865–2871. URL: https://doi.org/10.1093/bioinformatics/btp394.

[339] Jian Ye *et al.* "Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction". In: *BMC Bioinformatics* 13 (1 2012), p. 134. URL: https://doi.org/10.1186/1471-2105-13-134.

[340] Tamara J Blätte *et al.* "getITD for FLT3-ITD-based MRD monitoring in AML". In: *Leukemia* 33 (10 2019), pp. 2535–2539. URL: https://doi.org/10.1038/s41375-019-0483-z.

[341] Marcel Martin. "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *EMBnet.journal* 17.1 (2011), pp. 10–12. URL: https://journal.embnet.org/index.php/embnetjournal/article/view/200.

[342] Benjamin Kaminow, Dinar Yunusov, and Alexander Dobin. "STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data". In: *bioRxiv* (2021), p. 2021.05.05.442755. URL: https://www.biorxiv.org/content/biorxiv/early/2021/05/05/2021.05.05.442755.full.pdf.

[343] Adam Frankish *et al.* "GENCODE: reference annotation for the human and mouse genomes in 2023". In: *Nucleic Acids Research* 51.D1 (Nov. 2022), pp. D942–D949. URL: https://doi.org/10.1093/nar/gkac1071.

[344] Yuhan Hao *et al.* "Dictionary learning for integrative, multimodal and scalable single-cell analysis". In: *Nature Biotechnology* 42 (2 2024), pp. 293–304. URL: https://doi.org/10.1038/s41587-023-01767-y.

[345] Robert A Amezquita *et al.* "Orchestrating single-cell analysis with Bioconductor". In: *Nature Methods* 17 (2 2020), pp. 137–145. URL: https://doi.org/10.1038/s41592-019-0654-x.

[346] Anoop P Patel *et al.* "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma". In: *Science* 344 (6190 2014), pp. 1396–1401. URL: https://www.science.org/doi/abs/10.1126/science.1254257.

[347] Jack Kuipers *et al.* "Single-cell sequencing data reveal widespread recurrence and loss of mutational hits in the life histories of tumors". In: *Genome Research* 27 (11 2017), pp. 1885–1894. URL: http://genome.cshlp.org/content/27/11/1885.abstract.

[348] M Ryan Corces *et al.* "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues". In: *Nature Methods* 14 (10 2017), pp. 959–962. URL: https://doi.org/10.1038/nmeth.4396.

[349] Xi Chen *et al.* "A rapid and robust method for single cell chromatin accessibility profiling". In: *Nature Communications* 9 (1 2018), p. 5345. URL: https://doi.org/10.1038/s41467-018-07771-0.

[350] Wei Xu *et al.* "A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility". In: *Nature Protocols* 16 (8 2021), pp. 4084–4107. URL: https://doi.org/10.1038/s41596-021-00583-5.

[351] Daehwan Kim *et al.* "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype". In: *Nature Biotechnology* 37 (8 2019), pp. 907–915. URL: https://doi.org/10.1038/s41587-019-0201-4.

[352] Yong Zhang *et al.* "Model-based Analysis of ChIP-Seq (MACS)". In: *Genome Biology* 9 (9 2008), R137. URL: https://doi.org/10.1186/gb-2008-9-9-r137.

[353] Luis R Nassar *et al.* "The UCSC Genome Browser database: 2023 update". In: *Nucleic Acids Research* 51 (D1 2022), pp. D1188–D1195. URL: https://doi.org/10.1093/nar/gkac1072.

[354] Lance D Hentges *et al.* "LanceOtron: a deep learning peak caller for genome sequencing experiments". In: *Bioinformatics* 38 (18 2022), pp. 4255–4263. URL: https://doi.org/10.1093/bioinformatics/btac525.

[355] Jeffrey M Granja *et al.* "ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis". In: *Nature Genetics* 53 (3 2021), pp. 403–411. URL: https://doi.org/10.1038/s41588-021-00790-6.

[356] Darren A. Cusanovich *et al.* "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing". In: *Science* 348.6237 (2015), pp. 910–914. URL: https://www.science.org/doi/abs/10.1126/science.aab1601.

[357] David van Dijk *et al.* "Recovering Gene Interactions from Single-Cell Data Using Data Diffusion". In: *Cell* 174 (3 2018), 716–729.e27. URL: https://doi.org/10.1016/j.cell.2018.05.061.

[358] Dorin-Mirel Popescu *et al.* "Decoding human fetal liver haematopoiesis". In: *Nature* 574 (7778 2019), pp. 365–371. URL: https://doi.org/10.1038/s41586-019-1652-y.

[359] Noa Novershtern *et al.* "Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis". In: *Cell* 144 (2 2011), pp. 296–309. URL: https://doi.org/10.1016/j.cell.2011.01.004.

[360] Lu Chen *et al.* "Transcriptional diversity during lineage commitment of human blood progenitors". In: *Science* 345 (6204 2014), p. 1251033. URL: https://www.science.org/doi/abs/10.1126/science.1251033.

[361] Uku Raudvere *et al.* "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)". In: *Nucleic Acids Research* 47 (W1 2019), W191–W198. URL: https://doi.org/10.1093/nar/gkz369.

[362] Alicia N Schep *et al.* "chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data". In: *Nature Methods* 14 (10 2017), pp. 975–978. URL: https://doi.org/10.1038/nmeth.4401.

[363] Gennady Korotkevich *et al.* "Fast gene set enrichment analysis". In: *bioRxiv* (2021). URL: https://www.biorxiv.org/content/early/2021/02/01/060012.

[364] Aravind Subramanian *et al.* "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43 (2005), pp. 15545–15550. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0506580102.

[365] Tim Stuart *et al.* "Single-cell chromatin state analysis with Signac". In: *Nature Methods* 18 (11 2021), pp. 1333–1341. URL: https://doi.org/10.1038/s41592-021-01282-5.

[366] Jin He *et al.* "Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes". In: *Nature Cell Biology* 15 (4 2013), pp. 373–384. URL: https://doi.org/10.1038/ncb2702.

[367] Vincent van den Boom *et al.* "Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis". In: *Cell Reports* 14 (2 2016), pp. 332–346. URL: https://doi.org/10.1016/j.celrep.2015.12.034.

[368] Shanshan Pei *et al.* "Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in Patients with Acute Myeloid Leukemia". In: *Cancer Discovery* 10 (4 2020), pp. 536–551. URL: https://doi.org/10.1158/2159-8290.CD-19-0710.

[369] Gideon Bollag *et al.* "Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells". In: *Nature Genetics* 12 (2 1996), pp. 144–148. URL: https://doi.org/10.1038/ng0296-144.

[370] Arthur M Buchberg *et al.* "Sequence homology shared by neurofibromatosis type-1 gene and IRA-1 and IRA-2 negative regulators of the RAS cyclic AMP pathway". In: *Nature* 347 (6290 1990), pp. 291–294. URL: https://doi.org/10.1038/347291a0.

[371] Roymarie Ballester *et al.* "The NF1 locus encodes a protein functionally related to mammalian GAP and yeast IRA proteins". In: *Cell* 63 (4 1990), pp. 851–859. URL: https://www.sciencedirect.com/science/article/pii/0092867490901514.

[372] Alan List, Benjamin L Ebert, and Pierre Fenaux. "A decade of progress in myelodysplastic syndrome with chromosome 5q deletion". In: *Leukemia* 32 (7 2018), pp. 1493–1499. URL: https://doi.org/10.1038/s41375-018-0029-9.

[373] Sayantanee Dutta *et al.* "Acute Myeloid Leukemia with Isolated Trisomy 4 Is Characterized By Frequent Mutations of TET2 and a Distinct DNA Methylation Signature". In: *Blood* 140 (Supplement 1 2022), pp. 8912–8913. URL: https://doi.org/10.1182/blood-2022-168939.

[374] Diane Roulston *et al.* "Molecular Genetics of Myeloid Leukemia: Identification of the Commonly Deleted Segment of Chromosome 20". In: *Blood* 82 (11 1993), pp. 3424–3429. URL: https://www.sciencedirect.com/science/article/pii/S000649712067954X.

[375] Domenica Caramazza *et al.* "Chromosome 1 abnormalities in myeloid malignancies: a literature survey and karyotype–phenotype associations". In: *European Journal of Haematology* 84 (3 2010), pp. 191–200. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0609.2009.01392.x.

[376] Hussein A Abbas *et al.* "Clinical and molecular profiling of AML patients with chromosome 7 or 7q deletions in the context of TP53 alterations and venetoclax treatment". In: *Leukemia & Lymphoma* 63 (13 2022), pp. 3105–3116. URL: https://doi.org/10.1080/10428194.2022.2118533.

[377] Yueh-Chwen Hsu *et al.* "The distinct biological implications of Asxl1 mutation and its roles in leukemogenesis revealed by a knock-in mouse model". In: *Journal of Hematology & Oncology* 10 (1 2017), p. 139. URL: https://doi.org/10.1186/s13045-017-0508-x.

[378] Reina Nagase *et al.* "Expression of mutant Asxl1 perturbs hematopoiesis and promotes susceptibility to leukemic transformation". In: *Journal of Experimental Medicine* 215 (6 2018), pp. 1729–1747. URL: https://doi.org/10.1084/jem.20171151.

[379] Jason D Buenrostro *et al.* "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". In: *Nature Methods* 10 (12 2013), pp. 1213–1218. URL: https://doi.org/10.1038/nmeth.2688.

[380] Jason D Buenrostro *et al.* "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide". In: *Current Protocols in Molecular Biology* 109 (1 2015), pp. 21.29.1–21.29.9. URL: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471142727.mb2129s109.

[381] Ansuman T Satpathy *et al.* "Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion". In: *Nature Biotechnology* 37 (8 2019), pp. 925–936. URL: https://doi.org/10.1038/s41587-019-0206-z.

[382] Jason D Buenrostro *et al.* "Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation". In: *Cell* 173 (6 2018), 1535–1548.e16. URL: https://www.sciencedirect.com/science/article/pii/S009286741830446X.

[383] Sai Ma *et al.* "Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin". In: *Cell* 183 (4 2020), 1103–1116.e20. URL: https://www.sciencedirect.com/science/article/pii/S0092867420312538.

[384] Caleb A Lareau *et al.* "Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling". In: *Nature Biotechnology* 39 (4 2021), pp. 451–461. URL: https://doi.org/10.1038/s41587-020-0645-6.

[385] Leif S Ludwig *et al.* "Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics". In: *Cell* 176 (6 2019), 1325–1339.e22. URL: https://www.sciencedirect.com/science/article/pii/S0092867419300558.

[386] Chen Weng *et al.* "Deciphering cell states and genealogies of human haematopoiesis". In: *Nature* 627 (8003 2024), pp. 389–398. URL: https://doi.org/10.1038/s41586-024-07066-z.

[387] Michael Scherer *et al.* "Somatic epimutations enable single-cell lineage tracing in native hematopoiesis across the murine and human lifespan". In: *bioRxiv* (2024), p. 2024.04.01.587514. URL: https://www.biorxiv.org/content/biorxiv/early/2024/04/01/2024.04.01.587514.full.pdf.

[388] Franco Izzo *et al.* "Mapping genotypes to chromatin accessibility profiles in single cells". In: *Nature* 629 (8014 2024), pp. 1149–1157. URL: https://doi.org/10.1038/s41586-024-07388-y.

[389] John G Tate *et al.* "COSMIC: the Catalogue Of Somatic Mutations In Cancer". In: *Nucleic Acids Research* 47 (D1 2018), pp. D941–D947. URL: https://doi.org/10.1093/nar/gky1015.

[390] Courtney D DiNardo and Jorge E Cortes. "Mutations in AML: prognostic and therapeutic implications". In: *Hematology* 2016 (1 2016), pp. 348–355. URL: https://doi.org/10.1182/asheducation-2016.1.348.

[391] Sarah E Pierce, Jeffrey M Granja, and William J Greenleaf. "High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer". In: *Nature Communications* 12 (1 2021), p. 2969. URL: https://doi.org/10.1038/s41467-021-23213-w.

[392] Cizhong Jiang and B Franklin Pugh. "Nucleosome positioning and gene regulation: advances through genomics". In: *Nature Reviews Genetics* 10 (3 2009), pp. 161–172. URL: https://doi.org/10.1038/nrg2522.

[393] Kazuki Terada *et al.* "Usefulness of BCOR gene mutation as a prognostic factor in acute myeloid leukemia with intermediate cytogenetic prognosis". In: *Genes, Chromosomes and Cancer* 57 (8 2018), pp. 401–408. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/gcc.22542.

[394] A Greif Philipp *et al.* "RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes". In: *Haematologica* 97 (12 2012), pp. 1909–1915. URL: https://haematologica.org/article/view/6508.

[395] Jerónimo Bravo *et al.* "The leukemia-associated AML1 (Runx1)–CBFβ complex functions as a DNA-induced molecular clamp". In: *Nature Structural Biology* 8 (4 2001), pp. 371–378. URL: https://doi.org/10.1038/86264.

[396] Hanif Ullah *et al.* "In-silico probing of AML related RUNX1 cancer-associated missense mutations: Predicted relationships to DNA binding and drug interactions". In: *Frontiers in Molecular Biosciences* 9 (2022). URL: https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2022.981020.

[397] Woo-Joo Song *et al.* "Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia". In: *Nature Genetics* 23 (2 1999), pp. 166–175. URL: https://doi.org/10.1038/13793.

[398] Franco Izzo *et al.* "DNA methylation disruption reshapes the hematopoietic differentiation landscape". In: *Nature Genetics* 52 (4 2020), pp. 378–387. URL: https://doi.org/10.1038/s41588-020-0595-4.

[399] Ross M Okamura *et al.* "Redundant Regulation of T Cell Differentiation and TCRalpha Gene Expression by the Transcription Factors LEF-1 and TCF-1". In: *Immunity* 8 (1 1998), pp. 11–20. URL: https://doi.org/10.1016/S1074-7613(00)80454-9.

[400] Andrew M Intlekofer *et al.* "Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin". In: *Nature Immunology* 6 (12 2005), pp. 1236–1244. URL: https://doi.org/10.1038/ni1268.

[401] Michael A Paley *et al.* "Progenitor and Terminal Subsets of CD8+ T Cells Cooperate to Contain Chronic Viral Infection". In: *Science* 338 (6111 2012), pp. 1220–1225. URL: https://www.science.org/doi/abs/10.1126/science.1229620.

[402] Won Fen Wong *et al.* "Interplay of transcription factors in T-cell differentiation and function: the role of Runx". In: *Immunology* 132 (2 2011), pp. 157–164. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2567.2010.03381.x.

[403] Fiona C Brown *et al.* "MEF2C Phosphorylation Is Required for Chemotherapy Resistance in Acute Myeloid Leukemia". In: *Cancer Discovery* 8 (4 2018), pp. 478–497. URL: https://doi.org/10.1158/2159-8290.CD-17-1271.

[404] Maike Schwieger *et al.* "Homing and invasiveness of MLL/ENL leukemic cells is regulated by MEF2C". In: *Blood* 114 (12 2009), pp. 2476–2488. URL: https://doi.org/10.1182/blood-2008-05-158196.

[405] Lianzhong Zhao *et al.* "Transcription factor MEF2D is required for the maintenance of MLL-rearranged acute myeloid leukemia". In: *Blood Advances* 5 (22 2021), pp. 4727–4740. URL: https://doi.org/10.1182/bloodadvances.2021004469.

[406] Mathias Uhlén *et al.* "Tissue-based map of the human proteome". In: *Science* 347.6220 (2015), p. 1260419. URL: https://www.science.org/doi/abs/10.1126/science.1260419.

[407] Max Karlsson *et al.* "A single–cell type transcriptomics map of human tissues". In: *Science Advances* 7.31 (2021), eabh2169. URL: https://www.science.org/doi/abs/10.1126/sciadv.abh2169.

[408] Jason H Mendler *et al.* "RUNX1 Mutations Are Associated With Poor Outcome in Younger and Older Patients With Cytogenetically Normal Acute Myeloid Leukemia and With Distinct Gene and MicroRNA Expression Signatures". In: *Journal of Clinical Oncology* 30 (25 2012), pp. 3109–3118. URL: https://ascopubs.org/doi/abs/10.1200/JCO.2011.40.6652.

[409] Joseph D Khoury *et al.* "The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms". In: *Leukemia* 36 (7 2022), pp. 1703–1719. URL: https://doi.org/10.1038/s41375-022-01613-1.

[410] Daniel A Arber *et al.* "Classification of myeloid neoplasms/acute leukemia: Global perspectives and the international consensus classification approach". In: *American Journal of Hematology* 97 (5 2022), pp. 514–518. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajh.26503.

[411] Koichi Takahashi *et al.* "Integrative genomic analysis of adult mixed phenotype acute leukemia delineates lineage associated molecular subtypes". In: *Nature Communications* 9 (1 2018), p. 2670. URL: https://doi.org/10.1038/s41467-018-04924-z.

[412] Gabriele Merati *et al.* "Enrichment of Double RUNX1 Mutations in Acute Leukemias of Ambiguous Lineage". In: *Frontiers in Oncology* 11 (2021). URL: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.726637.

[413] Michael Heuser *et al.* "MN1 overexpression induces acute myeloid leukemia in mice and predicts ATRA resistance in patients with AML". In: *Blood* 110 (5 2007), pp. 1639–1647. URL: https://doi.org/10.1182/blood-2007-03-080523.

[414] Michael Heuser *et al.* "High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics". In: *Blood* 108 (12 2006), pp. 3898–3905. URL: https://doi.org/10.1182/blood-2006-04-014845.

[415] Michael Heuser *et al.* "Cell of Origin in AML: Susceptibility to MN1-Induced Transformation Is Regulated by the MEIS1/AbdB-like HOX Protein Complex". In: *Cancer Cell* 20 (1 2011), pp. 39–52. URL: https://doi.org/10.1016/j.ccr.2011.06.020.

[416] Masayuki Yamashita and Emmanuelle Passegué. "TNF-α Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration". In: *Cell Stem Cell* 25 (3 2019), 357–372.e7. URL: https://doi.org/10.1016/j.stem.2019.05.019.

[417] Kasper Jermiin Knudsen *et al.* "ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation". In: *Genes Development* 29.18 (2015), pp. 1915–1929. URL: http://genesdev.cshlp.org/content/29/18/1915.abstract.

[418] Zejuan Li *et al.* "PBX3 and MEIS1 Cooperate in Hematopoietic Cells to Drive Acute Myeloid Leukemias Characterized by a Core Transcriptome of the MLL-Rearranged Disease". In: *Cancer Research* 76 (3 2016), pp. 619–629. URL: https://doi.org/10.1158/0008-5472.CAN-15-1566.

[419] Zejuan Li *et al.* "PBX3 is an important cofactor of HOXA9 in leukemogenesis". In: *Blood* 121 (8 2013), pp. 1422–1431. URL: https://doi.org/10.1182/blood-2012-07-442004.

[420] Piu Wong *et al.* "Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential". In: *Genes Development* 21 (21 2007), pp. 2762–2774. URL: http://genesdev.cshlp.org/content/21/21/2762.abstract.

[421] Rampal Raajit and E Figueroa Maria. "Wilms tumor 1 mutations in the pathogenesis of acute myeloid leukemia". In: *Haematologica* 101 (6 2016), pp. 672–679. URL: https://haematologica.org/article/view/7732.

[422] Frederik Damm *et al.* "BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders". In: *Blood* 122 (18 2013), pp. 3169–3177. URL: https://doi.org/10.1182/blood-2012-11-469619.

[423] Paolo Sportoletti, Daniele Sorcini, and Brunangelo Falini. "BCOR gene alterations in hematologic diseases". In: *Blood* 138 (24 2021), pp. 2455–2468. URL: https://doi.org/10.1182/blood.2021010958.

[424] Madison J Kelly *et al.* "Bcor loss perturbs myeloid differentiation and promotes leukaemogenesis". In: *Nature Communications* 10 (1 2019), p. 1347. URL: https://doi.org/10.1038/s41467-019-09250-6.

[425] Q Cao *et al.* "BCOR regulates myeloid cell proliferation and differentiation". In: *Leukemia* 30 (5 2016), pp. 1155–1165. URL: https://doi.org/10.1038/leu.2016.2.

[426] Simone Tamburri *et al.* "Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression". In: *Molecular Cell* 77 (4 2020), 840–856.e5. URL: https://www.sciencedirect.com/science/article/pii/S1097276519308901.

[427] Miguel Vidal and Katarzina Starowicz. "Polycomb complexes PRC1 and their function in hematopoiesis". In: *Experimental Hematology* 48 (2017), pp. 12–31. URL: https://www.sciencedirect.com/science/article/pii/S0301472X16307743.

[428] Zheng Wang *et al.* "A Non-canonical BCOR-PRC1.1 Complex Represses Differentiation Programs in Human ESCs". In: *Cell Stem Cell* 22 (2 2018), 235–251.e9. URL: https://www.sciencedirect.com/science/article/pii/S1934590917305027.

[429] Sarah E Junco *et al.* "Structure of the Polycomb Group Protein PCGF1 in Complex with BCOR Reveals Basis for Binding Selectivity of PCGF Homologs". In: *Structure* 21 (4 2013), pp. 665–671. URL: https://www.sciencedirect.com/science/article/pii/S0969212613000531.

[430] Anca M Farcas *et al.* "KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands". In: *eLife* 1 (2012). Ed. by Kevin Struhl, e00205. URL: https://doi.org/10.7554/eLife.00205.

[431] Sarah J Wong *et al.* "KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1". In: *Structure* 24 (10 2016), pp. 1795–1801. URL: https://www.sciencedirect.com/science/article/pii/S0969212616302222.

[432] Eva J Schaefer *et al.* "BCOR and BCORL1 Mutations Drive Epigenetic Reprogramming and Oncogenic Signaling by Unlinking PRC1.1 from Target Genes". In: *Blood Cancer Discovery* 3 (2 2022), pp. 116–135. URL: https://doi.org/10.1158/2643-3230.BCD-21-0115.

[433] Kevin Oakley *et al.* "Setbp1 promotes the self-renewal of murine myeloid progenitors via activation of Hoxa9 and Hoxa10". In: *Blood* 119 (25 2012), pp. 6099–6108. URL: https://doi.org/10.1182/blood-2011-10-388710.

[434] John F Woolley *et al.* "INPP4B drives lysosome biogenesis to restrict leukemic stem cell differentiation and promote leukemogenesis". In: *bioRxiv* (2022), p. 2021.03.25.437029. URL: https://www.biorxiv.org/content/biorxiv/early/2022/04/11/2021.03.25.437029.full.pdf.

[435] Caterina Vitali *et al.* "SOCS2 Controls Proliferation and Stemness of Hematopoietic Cells under Stress Conditions and Its Deregulation Marks Unfavorable Acute Leukemias". In: *Cancer Research* 75 (11 2015), pp. 2387–2399. URL: https://doi.org/10.1158/0008-5472.CAN-14-3625.

[436] Robert W Georgantas III *et al.* "Microarray and Serial Analysis of Gene Expression Analyses Identify Known and Novel Transcripts Overexpressed in Hematopoietic Stem Cells". In: *Cancer Research* 64 (13 2004), pp. 4434–4441. URL: https://doi.org/10.1158/0008-5472.CAN-03-3247.

[437] Chi Huu Nguyen *et al.* "SOCS2 is part of a highly prognostic 4-gene signature in AML and promotes disease aggressiveness". In: *Scientific Reports* 9 (1 2019), p. 9139. URL: https://doi.org/10.1038/s41598-019-45579-0.

[438] I Dzneladze *et al.* "INPP4B overexpression is associated with poor clinical outcome and therapy resistance in acute myeloid leukemia". In: *Leukemia* 29 (7 2015), pp. 1485–1495. URL: https://doi.org/10.1038/leu.2015.51.

[439] Sewa Rijal *et al.* "Inositol polyphosphate 4-phosphatase II (INPP4B) is associated with chemoresistance and poor outcome in AML". In: *Blood* 125 (18 2015), pp. 2815–2824. URL: https://doi.org/10.1182/blood-2014-09-603555.

[440] Sabryn A Hamila *et al.* "The INPP4B paradox: Like PTEN, but different". In: *Advances in Biological Regulation* 82 (2021), p. 100817. URL: https://www.sciencedirect.com/science/article/pii/S2212492621000336.

[441] Mo Chen *et al.* "JMJD1C is required for the survival of acute myeloid leukemia by functioning as a coactivator for key transcription factors". In: *Genes Development* 29 (20 2015), pp. 2123–2139. URL: http://genesdev.cshlp.org/content/29/20/2123.abstract.

[442] Yi-Hung Ou *et al.* "The candidate tumor suppressor BTG3 is a transcriptional target of p53 that inhibits E2F1". In: *The EMBO Journal* 26 (17 2007), pp. 3968–3980. URL: https://www.embopress.org/doi/abs/10.1038/sj.emboj.7601825.

[443] Huidong Guo *et al.* "PBX3 is Essential for Leukemia Stem Cell Maintenance in MLL-Rearranged Leukemia". In: *Blood* 128 (22 2016), p. 2694. URL: https://www.sciencedirect.com/science/article/pii/S0006497119326953.

[444] Nan Zhu *et al.* "MLL-AF9– and HOXA9-mediated acute myeloid leukemia stem cell self-renewal requires JMJD1C". In: *The Journal of Clinical Investigation* 126 (3 2016), pp. 997–1011. URL: https://doi.org/10.1172/JCI82978.

[445] Garcia-Cuellar Maria-Paz *et al.* "Pbx3 and Meis1 cooperate through multiple mechanisms to support Hox-induced murine leukemia". In: *Haematologica* 100 (7 2015), pp. 905–913. URL: https://haematologica.org/article/view/7436.

[446] Sarah Naomi Olsen *et al.* "MLL::AF9 degradation induces rapid changes in transcriptional elongation and subsequent loss of an active chromatin landscape". In: *Molecular Cell* 82 (6 2022), 1140–1155.e11. URL: https://www.sciencedirect.com/science/article/pii/S1097276522001186.

[447] Aniruddha J Deshpande *et al.* "AF10 Regulates Progressive H3K79 Methylation and HOX Gene Expression in Diverse AML Subtypes". In: *Cancer Cell* 26 (6 2014), pp. 896–908. URL: https://doi.org/10.1016/j.ccell.2014.10.009.

[448] Daniel J L Coleman *et al.* "Gene regulatory network analysis predicts cooperating transcription factor regulons required for FLT3-ITD+ AML growth". In: *Cell Reports* 42 (12 2023), p. 113568. URL: https://www.sciencedirect.com/science/article/pii/S2211124723015802.

[449] R A Alharbi *et al.* "The role of HOX genes in normal hematopoiesis and acute leukemia". In: *Leukemia* 27 (5 2013), pp. 1000–1008. URL: https://doi.org/10.1038/leu.2012.356.

[450] H A Drabkin *et al.* "Quantitative HOX expression in chromosomally defined subsets of acute myelogenous leukemia". In: *Leukemia* 16 (2 2002), pp. 186–195. URL: https://doi.org/10.1038/sj.leu.2402354.

[451] T R Golub *et al.* "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring". In: *Science* 286 (5439 1999), pp. 531–537. URL: https://www.science.org/doi/abs/10.1126/science.286.5439.531.

[452] Scott A Armstrong *et al.* "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia". In: *Nature Genetics* 30 (1 2002), pp. 41–47. URL: https://doi.org/10.1038/ng765.

[453] Ashley P Ng *et al.* "Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice". In: *Blood* 118 (9 2011), pp. 2454–2461. URL: https://doi.org/10.1182/blood-2011-03-344739.

[454] Céline Moison *et al.* "HMGA2 expression defines a subset of human AML with immature transcriptional signature and vulnerability to G2/M inhibition". In: *Blood Advances* 6 (16 2022), pp. 4793–4806. URL: https://doi.org/10.1182/bloodadvances.2021005828.

[455] Praveen Kumar *et al.* "HMGA2 promotes long-term engraftment and myeloerythroid differentiation of human hematopoietic stem and progenitor cells". In: *Blood Advances* 3 (4 2019), pp. 681–691. URL: https://doi.org/10.1182/bloodadvances.2018023986.

[456] Evert Kroon *et al.* "Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b". In: *The EMBO Journal* 17 (13 1998), pp. 3714–3725. URL: https://www.embopress.org/doi/abs/10.1093/emboj/17.13.3714.

[457] Joerg Faber *et al.* "HOXA9 is required for survival in human MLL-rearranged acute leukemias". In: *Blood* 113 (11 2009), pp. 2375–2385. URL: https://doi.org/10.1182/blood-2007-09-113597.

[458] Cailin T Collins and Jay L Hess. "Deregulation of the HOXA9/MEIS1 axis in acute leukemia". In: *Current Opinion in Hematology* 23 (4 2016), pp. 354–361. URL: https://journals.lww.com/co-hematology/fulltext/2016/07000/deregulation_of_the_hoxa9_meis1_axis_in_acute.10.aspx.

[459] Jon Mar Björnsson *et al.* "Proliferation of primitive myeloid progenitors can be reversibly induced by HOXA10". In: *Blood* 98 (12 2001), pp. 3301–3308. URL: https://doi.org/10.1182/blood.V98.12.3301.

[460] Unnur Thorsteinsdottir *et al.* "Overexpression of HOXA10 in Murine Hematopoietic Cells Perturbs both Myeloid and Lymphoid Differentiation and Leads to Acute Myeloid Leukemia". In: *Molecular and Cellular Biology* 17 (1 1997), pp. 495–505. URL: https://doi.org/10.1128/MCB.17.1.495.

[461] William Hill *et al.* "Lung adenocarcinoma promotion by air pollutants". In: *Nature* 616 (7955 2023), pp. 159–167. URL: https://doi.org/10.1038/s41586-023-05874-3.

[462] Sung Choe *et al.* "Molecular mechanisms mediating relapse following ivosidenib monotherapy in IDH1-mutant relapsed or refractory AML". In: *Blood Advances* 4 (9 2020), pp. 1894–1905. URL: https://doi.org/10.1182/bloodadvances.2020001503.

[463] C D DiNardo *et al.* "Molecular patterns of response and treatment failure after frontline venetoclax combinations in older patients with AML". In: *Blood* 135 (11 2020), pp. 791–803. URL: https://doi.org/10.1182/blood.2019003988.

[464] Jun Ma *et al.* "Inhibition of Bcl-2 Synergistically Enhances the Antileukemic Activity of Midostaurin and Gilteritinib in Preclinical Models of FLT3-Mutated Acute Myeloid Leukemia". In: *Clinical Cancer Research* 25 (22 2019), pp. 6815–6826. URL: https://doi.org/10.1158/1078-0432.CCR-19-0832.

[465] Terrence N Wong *et al.* "Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia". In: *Nature* 518 (7540 2015), pp. 552–555. URL: https://doi.org/10.1038/nature13968.

[466] Kerryn Elliott and Erik Larsson. "Non-coding driver mutations in human cancer". In: *Nature Reviews Cancer* 21 (8 2021), pp. 500–509. URL: https://doi.org/10.1038/s41568-021-00371-z.

[467] Caroline J Watson *et al.* "The evolutionary dynamics and fitness landscape of clonal hematopoiesis". In: *Science* 367 (6485 2020), pp. 1449–1454. URL: https://www.science.org/doi/abs/10.1126/science.aay9333.

[468] Caroline J Watson and Jamie R Blundell. "Mutation rates and fitness consequences of mosaic chromosomal alterations in blood". In: *Nature Genetics* 55 (10 2023), pp. 1677–1685. URL: https://doi.org/10.1038/s41588-023-01490-z.

[469] Emily B Heikamp *et al.* "Epigenetic Resistance to Menin-MLL1 Inhibition Is Driven By Loss of the Non-Canonical Polycomb Repressive Complex 1.1 in NUP98-Rearranged AML". In: *Blood* 140 (Supplement 1 2022), pp. 95–96. URL: https://doi.org/10.1182/blood-2022-170165.

[470] Kai Yu *et al.* "Genomic Landscape of Patients with Germline RUNX1 Variants and Familial Platelet Disorder with Myeloid Malignancy". In: *bioRxiv* (2023), p. 2023.01.17.524290. URL: https://www.biorxiv.org/content/biorxiv/early/2023/01/17/2023.01.17.524290.full.pdf.

[471] Roel G W Verhaak *et al.* "Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance". In: *Blood* 106 (12 2005), pp. 3747–3754. URL: https://doi.org/10.1182/blood-2005-05-2168.

[472] Akira Honda *et al.* "Loss-of-function mutations in BCOR contribute to chemotherapy resistance in acute myeloid leukemia". In: *Experimental Hematology* 101-102 (2021), 42–48.e11. URL: https://www.sciencedirect.com/science/article/pii/S0301472X21002800.

[473] Teerapong Siriboonpiputtana *et al.* "Transcriptional memory of cells of origin overrides β-catenin requirement of MLL cancer stem cells". In: *The EMBO Journal* 36 (21 2017), pp. 3139–3155. URL: https://www.embopress.org/doi/abs/10.15252/embj.201797994.

[474] Bernd B Zeisig *et al.* "Functional reconstruction of human AML reveals stem cell origin and vulnerability of treatment-resistant MLL-rearranged leukemia". In: *Science Translational Medicine* 13 (582 2021), eabc4822. URL: https://www.science.org/doi/abs/10.1126/scitranslmed.abc4822.

[475] Indranil Singh *et al.* "Pre-existing stem cell heterogeneity dictates clonal responses to acquisition of cancer driver mutations". In: *bioRxiv* (2024). URL: https://www.biorxiv.org/content/early/2024/05/17/2024.05.14.594084.

[476] M Ryan Corces *et al.* "The chromatin accessibility landscape of primary human cancers". In: *Science* 362 (6413 2018), eaav1898. URL: https://www.science.org/doi/abs/10.1126/science.aav1898.

[477] Wei Xu *et al.* "ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells". In: *Nature Methods* 19 (10 2022), pp. 1243–1249. URL: https://doi.org/10.1038/s41592-022-01601-4.