Mechanisms of disease transformation in myeloproliferative neoplasms



Jennifer O'Sullivan

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2022

Abstract

Myeloproliferative neoplasms are heterogeneous clonal haematological malignancies where disease initiation occurs by acquisition of a somatic mutation in a haematopoietic stem cell. This mutation most often occurs in a janus kinase-signal transducer and activator of transcription signalling gene which drives constitutive janus kinase-signal transducer and activator of transcription signalling, with ensuing clonal expansion of downstream progenitor cells. Importantly, intra-tumoural cellular and molecular heterogeneity underscores the phenotypic heterogeneity of myeloproliferative neoplasms. Various factors drive disease progression but additional somatic mutations contribute significantly. Current myeloproliferative neoplasm therapies show limited effects on modifying disease biology and treatment resistance is common, associated with disease progression. Understanding whether distinct mutational profiles are associated with treatment resistance and examining the molecular signatures of mutant disease-driving haematopoietic stem and progenitor cells may reveal novel therapeutic targets and most importantly, lead to development of new treatments for these patients. Single-cell genomics approaches are ideally placed to achieve this through multi-omic technologies that can isolate mutant and non-mutant disease-driving stem and progenitor cells.

My overarching aims were to firstly to understand whether distinct mutational profiles were associated with treatment-resistant/intolerant myeloproliferative neoplasms and increased risk of disease progression and secondly to gain insights into the mechanisms driving transformation within a molecularly-defined myeloproliferative neoplasm subgroup through studying the disease-driving haematopoietic stem and progenitor cells. Study of two treatment-intolerant/resistant myeloproliferative neoplasm cohorts uncovered distinct molecular profiles. In hydroxycarbamide-resistant/intolerant essential thrombocythaemia, I observed increased *TP53* and *SF3B1* mutations in patients with myelofibrotic progression which importantly was not alleviated with ruxolitinib treatment. Analysis of a second, more advanced myeloproliferative neoplasm cohort of ruxolitinib-resistant/intolerant myelofibrosis patients showed a novel finding of enhanced RAS-pathway mutations.

To understand how *SF3B1* mutations collaborated with *JAK2*V617F mutations to accelerate myelofibrotic progression, I applied two single-cell multi-omic approaches. The first, 10X genomics, characterised the haematopoietic cellular landscape in *JAK2*V617F-*SF3B1*-mutated

myeloproliferative neoplasms and the second, TARGET-seq, paired mutation and gene expression information in the same cell to enable gene and splicing signature analysis of the mutated disease-driving and also non-mutant haematopoietic stem and progenitor cells.

Through 10X genomics analysis, I identified an abnormal expansion in earlier erythroid, megakaryocyte and most prominently eosinophil-basophil-mast cell progenitors in *JAK2*V617F-*SF3B1*-mutated haematopoietic stem and progenitor cells. I resolved the clonal architecture of *JAK2*V617F-*SF3B1*-mutated myeloproliferative neoplasms, observing that in the majority of cases, double mutant cells were present and that *SF3B1* was the founding event.

Uniquely, full-length transcriptome TARGET-seq made analysis of *SF3B1*-mutated aberrant splicing events possible. This uncovered novel biologically relevant aberrant splicing events in the *STAT1* gene. These events were a reduction in intron retention which would have an expected consequence of increased STAT1 activation. This analysis couples abnormal splicing with enhanced janus kinase-signal transducer and activator of transcription signalling to support a mechanism for an accelerated myeloproliferative neoplasm phenotype in *JAK2*V617F-*SF3B1*-mutated myeloproliferative neoplasms.

In conclusion, this analysis established *SF3B1* mutations in myeloproliferative neoplasms promote myelofibrotic disease progression and resolved the clonal architecture of *JAK2*V617F-*SF3B1*-mutated myeloproliferative neoplasms identifying abnormally expanded haematopoietic stem and progenitor cells populations. *SF3B1*-mutant specific splicing signatures were discovered that would be expected to augment janus kinase-signal transducer and activator of transcription signalling supporting mechanism by which *SF3B1* cooperates with *JAK2*V617F to promote an accelerated myeloproliferative neoplasm phenotype. The overarching aim is translating this new knowledge into patient benefit. Incorporation of spliceosome mutation information into clinical tools may improve risk stratification and specifically for *JAK2*V617F-*SF3B1*-mutated myeloproliferative neoplasms, a move towards precision medicine may be achievable through development of therapies targeting *STAT1*.

Acknowledgements

Firstly, I would like to thank my supervisors Professors Adam Mead, Claire Harrison and Beth Psaila for endless support, encouragement, kindness and guidance throughout. Thank you for giving me the opportunity to work and learn in such an inspiring and exciting scientific environment.

Next, this project would not have come to fruition without a number of key individuals. Thank you to Dr. Sally-Ann Clarke, of the WIMM flow facility, for her help with sorting and troubleshooting. Her expertise, kindness and patients during the madness of a big sort day was invaluable. I would like to especially thank Dr. Affaf Aliouat who gave so much of her expert knowledge, time and help with TARGET-seq experiments from assisting sorting to library preparation and did so with such kindness and good nature. Thank you to Sean, now Dr. Wei Wen!, a co-DPhil student who assisted with the computational splicing analysis in the project by applying a novel pipeline he set up and without which would not have been possible. His energy, wisdom and patience working collaboratively together on this has been extremely appreciated. I am grateful to Dr. Angela Hamblin for her expertise, input and guidance for the bulk mutational analysis in this work. I would like thank Dr. Alba Rodriguez-Meira for her help, support and expert knowledge throughout this project. I am grateful to Neil Ashley, ex single-cell facility manager at the WIMM, for all the times he rescued a robot from human error! His humour, patience and wisdom was always very much appreciated.

I would also like to thank all of my colleagues at the WIMM, particularly of the Mead and Psaila laboratories (current and past) for their help, teaching, support, company and laughter during this period. Thank you especially to Lauren, Christina, Rong, Nikos, Charlotte and Guanlin.

I am grateful to the funders of this project, the Medical Research Council, who awarded me a fellowship to pursue this work. Importantly, a thank you to the patients who so generously donate their tissue for research and without whom, none of this would be possible.

Finally, I would like to thank my family (my parents, brothers and sister) and friends who have been so patient and supportive during this time.

Declaration

Experiments and analysis in this thesis were designed and performed by myself with contributions of colleagues acknowledged in relevant sections. Parts of the work in this thesis has been previously published including first authored review articles[1,2] and a first authored research article[3]. Sections of the review articles including figures have been adapted and incorporated into the introduction chapter 1 (Sections 1-1.1, 1-1.3, 1-1.4, 1-2, 1-3, 1-4). Parts of the research article including figures have been adapted and included in chapter 3.

The following contributions were specifically made in this work and are acknowledged in relevant sections throughout:

1. Sequencing of bulk mutation data was performed by Oxford molecular diagnostics lab (Section 2-2).
2. Cytokine assay was performed by an external partner, CTI BioPharma (Section 2-3).
3. Single-cell 10X/CITE-seq libraries were prepared by Dr. Neil Ashley in the Single-Cell Facility in the MRC Weatherall Institute of Molecular Medicine (Section 2-6).
4. Flow cytometry experiments were conducted with assistance of Dr. Sally-Ann Clarke, of the WIMM flow facility (Section 2-5).
5. Bioinformatic analysis was performed by Wei Wen for TARGET-seq splicing analysis (Section 5-2.3) and generating Figures 2.12, 5.7, 5.14 - 5.18 and 5.20A-B.

Publications

1. **O'Sullivan JM**, Mead AJ, Psaila B. Single-cell methods in myeloproliferative neoplasms: old questions, new technologies. Blood. 2023; 141: 380-390 doi: 10.1182/blood.2021014668

2. **O'Sullivan J,** Psaila B. Lords of the Iron Rings: Coordinated Missplicing of TMEM14C and ABCB7 Drives Ring Sideroblast Formation in SF3B1-Mutant MDS. The Hematologist, Oct 2022, Volume 19, Issue 6 (American Society of Hematology official member newsletter)

3. Sternberg A, Boucher R, Coulthard HC, Raghavan M, Culligan D, Jackson A, Cargo C, Dennis M, Metzner M, **O'Sullivan J**, Moore R, Bowen D, Vyas P. Phase Ib study of eltrombopag and azacitidine in patients with high-risk myelodysplastic syndromes and related disorders (the ELASTIC study). Br J Haematol. 2022 Oct;199(2):222-229. doi: 10.1111/bjh.18389.

4. Roy A, Wang G, Iskander D, O'Byrne S, Elliott N, **O'Sullivan J**, Buck G, Heuston EF, Wen WX, Meira AR, Hua P, Karadimitris A, Mead AJ, Bodine DM, Roberts I, Psaila B, Thongjuea S. Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. Cell Rep. 2021 Sep 14;36(11):109698. doi: 10.1016/j.celrep.2021.109698.

5. Saunders J, Curto-Garcia N, Sriskandarajah P, **O'Sullivan J**, Woodley C, Asirvatham S, Campbell-Drew M, Mathias J, Ellis T, Baker N, Radia DH, Ali S, Kordasti S, Harrington P, de Lavallade H, McLornan DP, Harrison CN. Large Scale Internet-based Survey of Patients With a Myeloproliferative Neoplasm: Opinions and Experiences Regarding SARS-CoV-2 (COVID-19) Vaccination Strategies in 2021 Hemasphere. 2021 Jun 23;5(7):e609. doi: 10.1097/HS9.0000000000000609.

6. Salisbury RA, Curto-Garcia N, **O'Sullivan J**, Chen F, Polzella P, Godfrey AL, Russell J, Knapper S, Willan J, Frewin R, Joshi S, Arami S, Burns S, Teh CH, Wadelin F, Dhanapal J, Neelakantan P, Milojkovic D, Psaila B, Szydlo R, Francis S, Cargo C, Jain M, McGregor A, Wallis L, Duncombe A, Hussein H, Dyer P, Munro L, Bond L, McMullin MF, Somervaille TCP, Garg M, Sekhar M, Harrison C, Mead AJ, Innes AJ. Results of a national UK physician reported survey of COVID-19 infection in patients with a myeloproliferative neoplasm Leukemia. 2021 Feb 12:1-7. doi: 10.1038/s41375-021-01143-2

7. Rodriguez-Meira A, **O'Sullivan J,** Rahman H, Mead AJ. "TARGET-Seq: A Protocol For High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing". STAR Protoc. 2020 Oct 21;1(3):100125. doi: 10.1016/j.xpro.2020.100125.

8. Chowdhury O, **O'Sullivan J,** Barkas N, Wang G, Buck G, Hamblin A, Tefferi A, Al-Ali HK, Barosi G, Devos T, Gisslinger H, Jiang Q, Kiladjian JJ, Mesa R, Passamonti F, Ribrag V, Schiller G, Vannucchi AM, Zhou D, McMullin MF, Zhong J, Gale RP, Mead AJ; RESUME trialists. Spliceosome mutations are common in persons with myeloproliferative neoplasm-associated myelofibrosis with RBC-transfusion dependence

and correlate with response to pomalidomide. Leukemia. 2020. doi: 10.1038/s41375-020-0979-6

9. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, Allen HL, Sanchis-Juan A, Frontini M, Thys C, Stephens J, Mapeta R, Burren OS, Downes K, Haimel M, Tuna S, Deevi SVV, Aitman TJ, Bennett DL, Calleja P, Carss K, Caulfield MJ, Chinnery PF, Dixon PH, Gale DP, James R, Koziell A, Laffan MA, Levine AP, Maher ER, Markus HS, Morales J, Morrell NW, Mumford AD, Ormondroyd E, Rankin S, Rendon A, Richardson S, Roberts I, Roy NBA, Saleem MA, Smith KGC, Stark H, Tan RYY, Themistocleous AC, Thrasher AJ, Watkins H, Webster AR, Wilkins MR, Williamson C, Whitworth J, Humphray S, Bentley DR; **NIHR BioResource for the 100,000 Genomes Project**, Kingston N, Walker N, Bradley JR, Ashford S, Penkett CJ, Freson K, Stirrups KE, Raymond FL, Ouwehand WH. Whole-genome sequencing of patients with rare diseases in a national health system. Nature. 2020 Jul;583(7814):96-102. doi: 10.1038/s41586-020-2434-2

10. Gerds AT, Savona MR, Scott BL, Talpaz M, Egyed M, Harrison CN, Yacoub A, Vannucchi A, Mead AJ, Kiladjian JJ, **O'Sullivan J**, García-Gutiérrez V, Bose P, Rampal RK, Miller CB, Palmer J, Oh ST, Buckley SA, Mould DR, Ito K, Tyavanagimatt S, Smith JA, Roman-Torres K, Devineni S, Craig AR, Mascarenhas Determining the recommended dose of pacritinib: results from the PAC203 dose-finding trial in advanced myelofibrosis. Blood Adv. 2020 Nov 24;4(22):5825-5835.doi:10.1182/bloodadvances.2020003314

11. Psaila B, Wang G, Rodriguez-Meira A, Li R, Heuston EF, Murphy L, Yee D, Hitchcock IS, Sousos N, **O'Sullivan J,** Anderson S, Senis YA, Weinberg OK, Calicchio ML; NIH Intramural Sequencing Center, Iskander D, Royston D, Milojkovic D, Roberts I, Bodine DM, Thongjuea S, Mead AJ. Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets. Mol Cell. 2020 May 7;78(3):477-492.e8. doi: 10.1016/j.molcel.2020.04.008

12. **O'Sullivan JM**, Hamblin A, Yap C, F, Boucher R, Panchal A, Alimam A, Dreau H, Howard K, Ware P, Cross NCP, McMullin MF, Harrison CN, Mead AJ. The poor outcome in high molecular risk, hydroxycarbamide-resistant/intolerant ET is not ameliorated by ruxolitinib. Blood. 2019; 134: 2107-2011 doi: 10.1182/blood.2019001861

13. **O'Sullivan J,** Mead AJ. Clonal Selection and Mutation-Therapy Interaction in the Development of Therapy-Related Myeloid Neoplasms. The Hematologist, May-June 2019, Volume 16, Issue 3 (American Society of Hematology official member newsletter)

14. Desterro J, McLornan DP Desterro J, McLornan DP, Curto Garcia N, **O'Sullivan J**, Alimam S, Keohane C, Woodley C, Francis Y, Kordasti S, Radia DH, Harrison CN. Essential thrombocythaemia treated with recombinant interferon: 'real world' United Kingdom referral centre experience. Br J Haematol. 2019 doi.org/10.1111/bjh.15968

15. **O'Sullivan J,** Mead AJ. Heterogeneity in myeloproliferative neoplasms: Causes and consequences. Adv Biol Regul. 2019; 71: 55-68.

Table of Contents

Abstract 2
Acknowledgements 4
Declaration 5
Publications 6
Table of Contents 8
List of figures 11
List of tables 14
Commonly used abbreviations 15
Chapter 1: Introduction 19
1-1 Myeloproliferative neoplasms (MPNs) 19
1-1.1 Background 19
1-1.2 Clinical features of MPNs 19
1-1.3 Pathophysiology in MPNs 20
1-1.4 Therapeutic strategies in MPNs 25
1-2 Heterogeneity in myeloproliferative neoplasms (MPNs) 26
1-2.1 Inherited and germline factors 27
1-2.2 Target cell acquiring disease-driving mutation 29
1-2.3 Clonal hierarchies and order of mutation acquisition 30
1-2.4 The bone marrow niche 30
1-3 Disease transformation in MPNs 31
1-3.1 MPN transformation presentation 31
1-3.2 Clonal evolution in MPN 32
1-4 Single-cell methods and their application in MPNs 33
1-4.1 Single-cell genomics methods 33
1-4.2 Single-cell genomics technologies in MPN 34
1-4.3 New single-cell approaches 35
1-4.4 Integration of single-cell genomics into clinical pathways 36
1-5 Hypothesis and aims 37
Chapter 2: Methods 39
2-1 Patient banking and sample selection 39
2-1.1 High-risk ET patients enrolled onto MAJIC-ET for mutation profiling 39
2-1.2 High-risk MF patients enrolled onto PAC203 study 39
2-1.3 Patients within distinct genetic MPN subgroup for single-cell analyses 39
2-2 Bulk mutation analysis 39
2-3 Cytokine assay in a high-risk MF cohort 40
2-4 Antibody-derived tag (ADT) selection for CITE-seq 41
2-5 Fluorescence-activated cell sorting (FACS) and single-cell isolation of HSPCs 43
2-5.1 Sample preparation and staining 43

2-5.2 10X genomics/CITE-seq FACS 46
2-5.3 TARGET-seq FACS 47

2-6 Single-cell methodology I: CITE-seq/10X Chromium (droplet-based) 49

2-7 Single-cell methodology II: TARGET-seq (plate-based) 51
2-7.1 Lysis buffer preparation 51
2-7.2 Single-cell cDNA synthesis 52
2-7.3 High-throughput single-cell genotyping 54
2-7.4 Single-cell full-length transcriptomics 59

2-8 Statistical analysis of clinico-molecular datasets 65

2-9 Bioinformatic analysis of 10X/CITE-seq data 66
2-9.1 Droplet-based single-cell RNA-sequencing and CITE-seq pre-processing 66
2-9.2 Donor demultiplexing, quality control and normalisation 66
2-9.3 Dimensionality reduction, data integration and clustering of 10X genomics data 68
2-9.4 Lineage trajectory analysis 69
2-9.5 VarTrix 69
2-9.6 Differential abundance analysis 70
2-9.7 Differential gene expression analysis 70
2-9.8 Regulatory network analysis 70
2-9.9 Single-cell genotyping analysis of plate-based 71
2-9.10 Reconstruction of clonal hierarchies in TARGET-seq genotyping data 71
2-9.11 TARGET-seq full-length transcriptomic pre-processing and analysis 72
2-9.12 Splicing event detection and quantification 73
2-9.13 Differential splicing and gene expression analysis 74

Chapter 3: The molecular landscape of high-risk MPN 75

3-1 Introduction 75
3-1.1 Hypothesis and aims 76

3-2 Results 77
3-2.1 The mutation profile of high-risk HC-RES/INT ET patients is distinct from low-risk ET 77
3-2.2 Correlation of mutation profiles with clinical responses and outcomes 81
3-2.3 Longitudinal mutation analysis in MAJIC-ET patients 85
3-2.4 Distinct mutational profile in RUX-RES/INT MF patients treated with JAK2 inhibition 88
3-2.5 *SF3B1* mutations in MPN 93

3-3 Discussion 94

Chapter 4: The molecular signatures associated with disease transformation in *JAK2*V617F-*SF3B1*-mutated MPN 98

4-1 Introduction 98
4-1.1 RNA splicing 98
4-1.2 RNA splicing is disrupted in cancer 100
4-1.3 *SF3B1* is recurrently mutated across cancer 100
4-1.4 *SF3B1* mutations in myeloid neoplasms 102
4-1.5 Hypothesis and aims 103

4-2 Results 103
4-2.1 Baseline characteristics of the dataset 103
4-2.2 Quality control and donor demultiplexing 105
4-2.3 CITE-seq 107
4-2.4 Data integration 108

4-2.5 Characterising HSPC and mature lineage compartments 110
4-2.6 Dysregulated gene signatures in JS+ HSPCs 113
4-2.7 Innate and adaptive immune system in JS+ MPN 121
4-2.8 Regulatory gene network analysis 124

4-3 Discussion 126

Chapter 5: Mutant-specific molecular signatures and aberrant splicing events in HSPCs of *JAK2*V617F-*SF3B1*-mutated MPN 129

5-1 Introduction 129
5-1.1 Hypothesis 132
5-1.2 Aims 133

5-2 Results 133
5-2.1 Clonal hierarchy and heterogeneity of JS+ MPN 133
5-2.2 Transcriptomic profiles of JS+ MPN 138
5-2.3 Aberrant splicing events in JS+ MPN 146

5-3 Discussion 152

Chapter 6: Discussion 155

6-1 Outcomes of this research 155

6-2 Limitations of this research 157

6-3 Future research 158

Bibliography 162

Appendix 1 184

Appendix 2 184

Appendix 3 184

Appendix 4 185

Appendix 5 185

List of figures

Figure 1.1. Haematopoietic hierarchy.
Figure 1.2. Schema of aberrant haematopoietic expansions in MPN.
Figure 1.3. Cytokine receptor activation in MPNs.
Figure 1.4. Factors influencing MPN disease heterogeneity.
Figure 1.5. Single-cell technologies.
Figure 1.6. Intra-patient single-cell multi-omic analysis of mutant and non-mutant hematopoietic stem and progenitor cells (HSPC).
Figure 1.7. Application of single-cell (SC) technologies in clinical trials**.**
Figure 2.1. Schema of ADT construct.
Figure 2.2. Schema for 10X genomics/CITE-seq and TARGET-seq sorting, library preparation and sequencing.
Figure 2.3. Populations sorted at enrichment.
Figure 2.4. Donor mixing strategy for 10X genomics/CITE-seq.
Figure 2.5. 384-well single-cell sorting.
Figure 2.6. Exomic, ADT and HTO library quality traces.
Figure 2.7. Dilution and separate genotyping and transcriptome stock plates.
Figure 2.8. Schema of genotyping library preparation**.**
Figure 2.9. Genotyping stock plate transferring to genotyping-PCR1 reaction plate layouts.
Figure 2.10. Genotyping-PCR2.
Figure 2.11. Tapestation traces of TARGET-seq genotyping library QC.
Figure 2.12. Downsampling analysis for splice junction analysis.
Figure 2.13. Illustrative schema of 384-well plate reformatting from source cDNA plates to destination reformatted plate.
Figure 2.14. Schema of full-length tagmentation method.
Figure 2.15. i5 and i7 index stock Deep-Well plate layouts.
Figure 3.1. Baseline mutational analysis and correlation with clinical characteristics and outcomes.
Figure 3.2. Kaplan-Meier curves of transformation-free survival (TFS) stratified by mutational statuses with survival estimates, reported at 4-years.
Figure 3.3. MAJIC-ET longitudinal mutational analysis.

Figure 3.4. Mutational landscape and clinical correlations in the PAC203 cohort.
Figure 3.5. Clinico-genomic characteristics of *SF3B1* mutations in MPN.
Figure 4.1. Schematic outline demonstrating alternative splicing of exons within single gene can generate multiple proteins.
Figure 4.2. U2-dependent (major) spliceosome.
Figure 4.3. Lollipop diagram of SF3B1.
Figure 4.4. 10X data donor demultiplexing and quality control.
Figure 4.5. Data integration and cell cluster annotation.
Figure 4.6. Characterisation of HSPC populations in the 10X dataset.
Figure 4.7. Lineage differentiation trajectories in 10X HSPCs**.**
Figure 4.8. Differential gene expression analysis in 10X HSPCs.
Figure 4.9. Gene expression of specific genes in 10X HSPCs.
Figure 4.10. Differential gene expression analysis in 10X HSC/MPP.
Figure 4.11. Differential gene expression analysis in 10X ErP.
Figure 4.12. Differentially expressed ADT markers in 10X HSPCs.
Figure 4.13. Mature cell population clustering.
Figure 4.14. Differential gene expression analysis in immune cell compartments.
Figure 4.15. Differential gene expression analysis in monocytes.
Figure 4.16. Regulatory networks in 10X HSPC populations.
Figure 5.1. Alternative splicing events.
Figure 5.2. Schema of *SF3B1*mut-associated A3SS.
Figure 5.3. Clonal composition of JS+ MPN samples.
Figure 5.4. Clonal composition of J+ MPN samples.
Figure 5.5. Integrated TARGET-seq genotyping and HSPC immunophenotype of JS+ HSPCs.
Figure 5.6. Integrated TARGET-seq genotyping and HSPC immunophenotype of J+ HSPCs.
Figure 5.7. TARGET-seq quality control metrics.
Figure 5.8. UMAP plots of TARGET-seq HSPCs merged without integration.
Figure 5.9. Analysis of TARGET-seq HSPC clusters.
Figure 5.10. Projection of TARGET-seq onto 10X HSPC populations.
Figure 5.11. Differential gene expression analysis of TARGET-seq *SF3B1*mut vs *SF3B1*wt HSPCs.
Figure 5.12. Differential gene expression analysis of TARGET-seq intra-patient mutant vs. WT HSPCs.

Figure 5.13. TARGET-seq WT HSPC analysis.

Figure 5.14. Global splicing events associated with *SF3B1*mut in the TARGET-seq dataset.

Figure 5.15. Identification of previously validated *SF3B1*mut-associated A3SS events.

Figure 5.16. Differentially splicing events (DSE) in *SF3B1*mut HSPCs.

Figure 5.17. Top differentially spliced events comparing inter- and intra-patient results.

Figure 5.18. *STAT1* ASE visualisation.

Figure 5.19. Schema of intron retention and effect on gene expression.

Figure 5.20. *STAT1* differential splicing and gene expression in TARGET-seq and 10X data.

Figure 6. Proposed mechanism by which *SF3B1* mutations accelerate fibrosis development in *JAK2*V617F MPN.

List of tables

Table 2.1. Totalseq cite-seq antibody markers included.

Table 2.2. HSPC immunophenotype panel used for FACS.

Table 2.3. ADT and HTO PCR reactions.

Table 2.4. Lysis buffer mastermix.

Table 2.5. RT mastermix.

Table 2.6. RT thermocycler program.

Table 2.7. PCR mastermix.

Table 2.8. PCR thermocycler program.

Table 2.9. Genotyping-PCR1 mastermix.

Table 2.10. Genotyping-PCR1 thermocycler program.

Table 2.11. Genotyping-PCR2 mastermix.

Table 2.12. Genotyping-PCR2 thermocycler program.

Table 2.13. Nextera XT Tagmentation reagents.

Table 2.14. Tagmentation thermocycler program.

Table 2.15. Indexing PCR thermocycler program.

Table 3.1. Baseline clinical characteristics in patients with and without non-MPN driver mutations (NDM).

Table 3.2. Logistic regression predicting the influence of non-MPN driver mutations (NDM) on clinical outcomes.

Table 3.3. *RAS*-mutated patient baseline clinical and mutation characteristics.

Table 4.1. Summary of patients and healthy controls used for 10X/CITE-seq and TARGET-seq experiments.

Commonly used abbreviations

ASE: Alternative splicing events
A3SS : Alternative 3' Splice Sites
A5SS : Alternative 5' Splice Sites
ADT: Antibody-derived Tag
AML : Acute Myeloid Leukaemia
ASXL1: Additional Sex Combs-Like 1
ATM: Amplicon Tagment Mix
BCL: Binary Base Call
B2M: Beta-2 Microglobulin
BDNF: Brain Derived Neurotrophic Factor
BM: Bone Marrow
Bp : base pair
BSA: Bovine Serum Albumin
CALR: calreticulin
CBL: Casitas B-lineage Lymphoma
CD: cluster of differentiation
CHIP : Clonal Haematopoiesis of Indeterminate Potential
CITE-seq: Cellular Indexing of Transcriptomes and Epitopes by Sequencing
CML : Chronic Myeloid Leukaemia
CMP : Common Myeloid Progenitor
CRP: C-reactive protein
Del: deletion
EBM: Eosinophil-Basophil-Mast cell
EPOR : Erythropoietin Receptor
ERCC : External RNA Controls Consortium
ET : Essential Thrombocythemia
EZH2 : Enhancer of Zeste Homolog 2
FACS: Fluorescent Activated Cell Sorting
FCS: Fetal Calf Serum
FDG: Force Directed Graph
FDA : Food and Drug Administration
FTH1: Ferritin Heavy Chain 1
FMO : Fluorescence Minus One
G-CSF-R : Granulocyte Colony Stimulating Factor Receptor
gDNA: genomic Deoxyribonucleic Acid
GH: Growth Hormone
GMCSF: Granulocyte Macrophage Colony Stimulating Factor Receptor
GMP : Granulocyte-Monocyte Progenitor
GSEA : Gene-set Enrichment Analysis
HIF1A: Hypoxia-Inducible Factor 1-Alpha
HSCs : Haematopoietic Stem Cells
HMR: High molecular risk
HSPCs : Haematopoietic Stem and Progenitor Cells
HC: Hydroxycarbamide
HTO: Hashing tagged oligonucleotides
ICAM1: Intercellular Adhesion Molecule 1

IFN-α : Interferon Alpha
IFN-γ: Interferon Gamma
IFNGR2: Interferon Gamma Receptor 2
IL-1ALP: Interleukin 1 alpha
IL-1-B: IL-1RA: interleukin 1 receptor alpha
IL-2: interleukin 2
IL-3: interleukin 3
IL-5: interleukin 4
IL-6: interleukin 5
IL-6: interleukin 6
IL-7: interleukin 7
IL-8: interleukin 8
IL-10: interleukin 8
IL-12P70: interleukin 12
IL-12P40: interleukin 8
IL-15: interleukin 15
IL-17: interleukin 17
HT1 : Hybridization Buffer 1
HTO: Hash-Tag Oligonucleotide
J+: *JAK2*V617F-mutated
JAK : Janus Activated Kinases
JAK2: Janus Kinase 2
*JAK2*V617Fmut: *JAK2*V617F mutant
JH1 : JAK Homology 1
JH2 : JAK Homology 2
JS+: *JAK2*V617F-*SF3B1*-mutated
KNN : K-nearest Neighbourhood
LNA : Locked Nucleic Acid
LMPP : Lymphoid-Primed Multipotent Progenitors
LOH: Loss of Heterozygosity
LOX1: Lectin-like oxLDL [oxidized low-density lipoprotein] receptor 1
LPA: Lipoprotein A
MAPK : Mitogen Activated Protein Kinase
MCP1: Monocyte chemoattractant protein-1
MDS : Myelodysplastic Syndrome
MEP : Megakaryocyte-Erythroid Progenitor MF : Myelofibrosis
MkP: Megakaryocyte progenitor
MHC: Major Histocompatibility Complex
MIP1AL: Macrophage inflammatory protein 1-alpha
MIP1BE: Macrophage inflammatory protein 1-beta
MMP3: Matrix Metallopeptidase 3
MMP9: Matrix Metallopeptidase 9
MNCs : Mononuclear cells
MPL: myeloproliferative leukaemia protein
MPN : Myeloproliferative neoplasms
MPP : Multipotent Progenitors
mRNA: messenger Ribonucleic Acid
MSigDB : Molecular Signatures Database
MXE : Mutually Exclusive Exons

NDM: non-MPN driver mutations
NFE2 : NF-E2 45kDa protein
NGS : Next Generation Sequencing
NTPROBN: N-terminal pro-brain natriuretic peptide
PB: peripheral blood
PBS : Phosphate Buffered Saline
PCA: Principal Component Analysis
PCR: Polymerase Chain Reaction
PI3K : Phosphoinositide 3-kinase
PPP1R15A: Protein Phosphatase 1 Regulatory Subunit 15A
PPP2R5A: Protein Phosphatase 2 Regulatory Subunit 5A
PSI: Percent Spliced In
PV: Polycythaemia Vera
QC: Quality Control
RAS: rat sarcoma
RI: Retained Inron
RNA-seq: Ribonucleic Acid Sequencing
RPCA: Reciprocal Principal Component Analysis
RT: Retrotranscription
RUX: Ruxolitinib
SAP: serum amyloid P component
sc-RNA-seq: single cell Ribonucleic Acid Sequencing
SD: Standard Deviation
SE: Skipped Exon
SF3B1: Splicing factor 3B subunit 1
*SF3B1*mut*: SF3B1* mutant
SMART : Switching Mechanism at the 5'-end of the RNA transcript
SNP: Single Nucleotide Polymorphism
snRNP: small nuclear ribonucleoprotein
STAR: Spliced Transcripts Alignment to Reference
STAT: Signal Transducers and Activators of Transcription
SRSF2: Serine And Arginine Rich Splicing Factor 2
TBG: T4-binding globulin
TET2: Tet methylcytosine dioxygenase 2
TGFβ : Transforming growth factor beta
TM: Thrombomodulin
TN : Triple Negative
TNFα : Tumour Necrosis Factor Alpha
TNFβ : Tumour Necrosis Factor Beta
TP53: Tumour Protein 53
tSNE: T-distributed stochastic neighbourhood
TSH: Thyroid Stimulating Hormone
TSO: Template Switching Oligonucleotides
U2AF1: U2 Small Nuclear RNA Auxiliary Factor 1
UMAP: Uniform Manifold Approximation and Projection
UMI : Unique Molecular Identifier
VAF: Variant Allele Frequency
VEGF: vascular endothelial growth factor
VWF: Von Willebrand Factor

WHO: World Health Organisation
WNN: Weighted Nearest Neighbor
WT: Wild-Type
ZRSR2: Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2

Chapter 1: Introduction

1-1 Myeloproliferative neoplasms (MPNs)

1-1.1 Background

Philadelphia chromosome negative myeloproliferative neoplasms (MPNs) are a heterogeneous group of haematological malignancies[4,5] typified by increased proliferation of one or all myeloid blood cell lineages. The disease is initiated by acquisition of a somatic mutation in a JAK-STAT (janus kinase-signal transducer and activator of transcription) signalling gene at the level of the haematopoietic stem cell (HSC)[6,7], which causes constitutive increased JAK-STAT signalling[8,9], with ensuing clonal expansion of downstream progenitor and mature blood cells. The most prevalent mutation driving MPN development is a single nucleotide base change in exon 14 of the *JAK2* gene, *JAK2*V617F mutation[10,11], followed by a deletion or insertion in the calreticulin (*CALR)* gene[12,13] and rarely a single nucleotide base change in the myeloproliferative leukaemia virus (*MPL)* gene[14]. These MPN driver mutations are mutually exclusive in the majority of cases. Additional somatic mutations in non-JAK-STAT signalling genes (e.g. epigenetic [*TET2*, *DNMT3A*], splicing [*SF3B1*, *ZRSR2*, *SRSF2*], metabolic [IDH1, IDH2] and other signalling pathway [*KRAS*, *NRAS*, *TP53*] genes) are present in MPN cases but more often at more advanced disease stages. In some cases, these mutations portent clonal evolution and progression to accelerated or blast phase disease. MPNs are excellent models of cancer evolution due to the long disease latency period, well-defined disease-driving mutations and ease of access for sampling throughout the disease course. While many scientific questions about MPN biology have remained the same for decades, the ability to answer these questions has advanced tremendously due to considerable technological advances.

1-1.2 Clinical features of MPNs

MPNs are categorised based on a combination of clinical and laboratory parameters, including mutation status, blood counts and histology into essential thrombocythaemia (ET), polycythaemia rubra vera (PV) and myelofibrosis (MF)[4,15]. Diagnosis of an MPN is established using an integrated approach excluding secondary causes of elevated blood counts (“cytosis”) and fulfilling WHO diagnostic criteria[16]. More recently, MPN diagnostic classifications was expanded to include

an additional earlier pre-fibrotic MF stage ("Pre-MF")[4,16]. ET and PV are earlier chronic phase MPNs. ET and PV have an estimated incidence of ~1 per 100,000[17] each and a median age at presentation of 68 and 65 years[18] respectively. ET has a preponderance in women[19,20] but PV is more common in men.

Patients with ET typically present with an isolated elevated platelet count and patients with PV most often present with an elevated haemocrit or haemoglobin level, although all myeloid lineages may be increased. For most patients, ET and PV follow a relatively benign course and in fact, many patients are diagnosed incidentally. Nevertheless, there is a notable impact on quality of life with risk of clinical sequelae; an augmented risk of thrombosis (both arterial and venous) or haemorrhage and heightened mortality when compared to healthy age-matched controls[21]. Increasingly, disease-related symptoms and their impact upon quality of life separate from these clinical events have also been recognised[22].

Bone marrow fibrosis, osteosclerosis and often massive splenomegaly due to extramedullary haematopoiesis are hallmarks of MF. Clinical consequences include cytopenias and transfusion dependence, infections, severe constitutional symptoms and premature mortality. All MPNs are preleukaemic conditions, with a risk of leukaemic transformation of 10-20% after 10 years for MF and much lower for ET and PV at 1-2% and 2-4% after 15 years respectively[23].

1-1.3 Pathophysiology in MPNs

i) Cell of origin: haematopoietic stem cell (HSC)

Haematopoiesis is a process that produces and maintains blood cells. In the classical model, HSCs are rare self-renewing cells differentiate in a step wise manner to more committed lineage progenitors without self-renewal capacity and ultimately to mature blood cells (Figure 1.1A). This haematopoietic hierarchy has been challenged in recent years through single-cell approaches[24] in both human and murine models but the conclusive findings have been derived from murine models[25]. These can be summarised by the following key concepts: (1) differentiation occurs along a continuum rather than in distinct stages, (2) megakaryocyte (MkP), erythroid (ErP), basophil, mast, eosinophil (EBM) progenitors differentiate more closely than previously appreciated and branch separately from myeloid and lymphoid progenitors and (3)

megakaryocyte differentiation can take place directly from HSC, potentially triggered by certain scenarios, bypassing intermediate progenitor stages (Figure 1.1B).

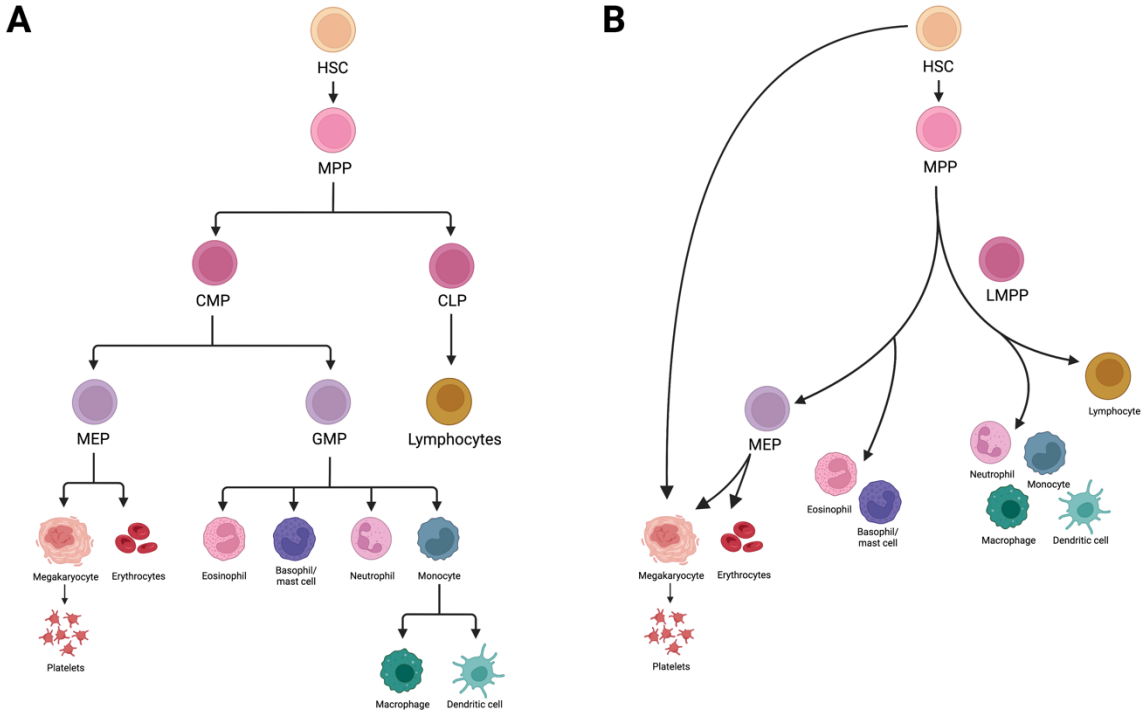


Figure 1.1. Haematopoietic hierarchy. (A) Classical model with stepwise transitions from the HSC to mature lineages. (B) New proposed model with differentiation along a continuum, a direct path for megakaryopoiesis from the HSC and closer differentiation trajectories for MkP, ErP and EBM progenitors.

Evidence supporting that the MPN cell of origin is likely to be a long term (LT)-HSC, situated at the apex of the haematopoietic hierarchy, is derived from mouse models[25]. Moreover, both *JAK2*V617F[6] and *CALR*[26] mutations are present at the immunophenotype-defined HSC level in patient samples of all MPN subtypes and furthermore *JAK2*V617F is detected in mature blood cells of multiple different lineages (myeloid and lymphoid), supporting a stem cell origin[25,27]. In *CALR*-mutated patients, a higher proportion of HSCs carry the mutation than *JAK2*V617F MPN (40% vs 8%)[28]. Yet, similarly to *JAK2*V617F, the *CALR*-mutated HSC has not demonstrated increased self-renewal capacity over wild-type (WT) HSCs in competitive murine transplant experiments[29,30]. In the xenograft setting, *JAK2*V617F-mutated human HSCs have been shown to result in increased mature myeloid output from the progenitors but with no increase in self-renewal or proliferation ability over WT HSCs[31]. In conditional knock-in mice, it has been reported that *JAK2*V617F may even potentially reduce the self-renewal capabilities of HSCs[32]. Nevertheless, disease-initiating potential appears to be confined to the HSC as transplantation of progenitor cells has hitherto failed to recapitulate an MPN phenotype in murine models[33]. In ET,

clonal expansion of megakaryocytes produces an excess of platelets (Figure 1.2A). Although an increase in erythroid progenitors dominates in PV, all myeloid are typically expanded (Figure 1.2B). In MF, heterogenous progenitors expand but megakaryocyte expansion is prominent and these cells interact with the bone marrow niche releasing pro-inflammatory and pro-fibrotic factors to develop fibrosis (Figure 1.2C).

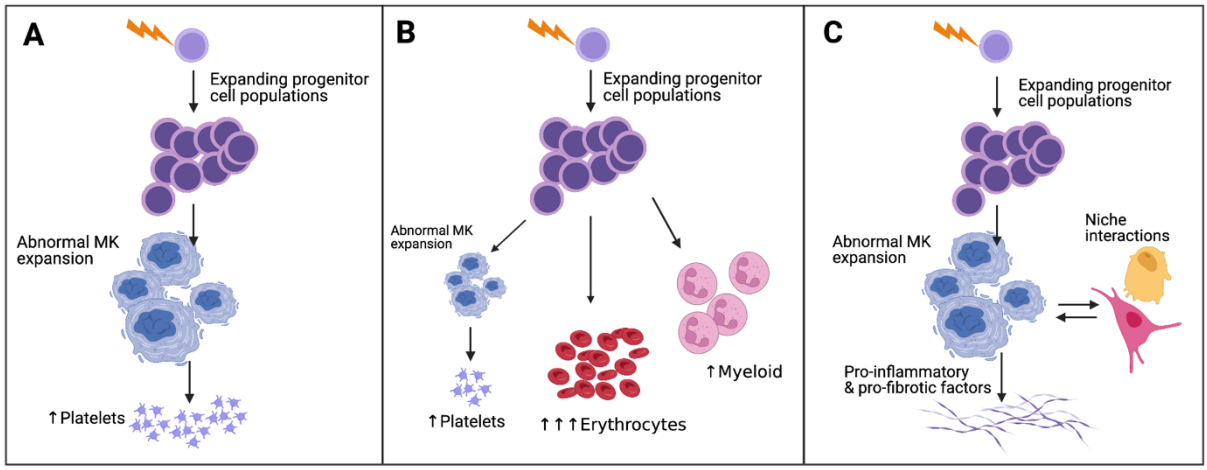


Figure 1.2 Schema of aberrant haematopoietic expansions in MPN. Disease-initiating mutation acquired by a HSC highlighted by orange lightning symbol. (A) MK expansion in ET. (B) Erythrocyte predominant expansion in PV. (C) Outline of abnormal MK expansion in MF with MK-niche interactions (osteoblast in yellow, stromal cell in pink) and fibrosis development. Adapted from O'Sullivan et al. *Blood* 2022[2].

ii) Disease-initiating mutations in genes of the JAK-STAT signalling pathway

Mutations in three genes, *JAK2*, *CALR*, and *MPL* account for majority of JAK-STAT signalling driver mutations in MPN. *JAK2*V617F, the first described and most prevalent disease-causing mutation in all MPNs, is a single nucleotide alteration in exon 14 of *JAK2* causing unrestrained upregulation in JAK-STAT signalling and downstream increased gene transcription and protein expression[10,11]. JAK-STAT signalling is enhances in all MPNs including cases without a detected driver mutation[8]. *JAK2*V617F is identified in >95% mutations of PV cases and ~50% of ET and MF cases. Mutations are also rarely described in exon 12 of *JAK2* explaining 1-2% of cases of PV[34].

One third of ET cases carry a mutation in exon 9 of the *CALR* gene which encodes for calreticulin, a chaperone protein with many functions including calcium homeostasis regulation[12,13]. Type 1 *CALR* mutations, a 52-base pair deletion (bpd), are the most common with type 2 mutations (5-

bp insertion) occurring in approximately 20%[35]. *CALR* mutations all cause a 1-bp frameshift mutation changing the negatively charged C-terminus to a mutant positively-charged terminus with neomorphic activity of upregulated JAK-STAT signalling through binding to the TPO receptor[36,37] but not the erythropoietin receptor. This receptor specificity explains why *CALR* mutations produce an ET or MF but never a PV phenotype.

The *MPL* gene encodes for the TPO receptor. Gain-of-function missense mutations in this gene[38] are reported in a minority of MPN cases: ~1% and 5% of ET[39] and MF[40]. These more often occur in older patients, never produce a PV phenotype and have been correlated with a heightened risk of fibrotic transformation[41].

Patients where no JAK-STAT driver mutation is identified are denoted "triple-negative" (TN) occurring in ~10% of MPNs[42]. However, some MPN cases may be misclassified as TN due to low-level driver mutations that fall below the threshold for detection by clinical diagnostic laboratories. An analysis of a TN ET cohort applied a deep sequencing approach detecting JAK-STAT signalling mutations in 22% (10/46) of patients. TN ET tends to have the most favourable prognosis[43] most often occurring in younger women. This is in stark contrast to TN MF where almost 90% have a non-JAK-STAT signalling mutation[44], the outcome is poor with limited survival as compared with driver-mutated disease[21]. This suggests triple-negative ET and PMF have a disparate disease biology.

iii) Canonical effects of constitutive JAK-STAT signalling pathway activation

The JAK2 pseudokinase domain (JH2) under normal circumstances acts as a negative regulator of the JH1 kinase domain which is the main domain for enzymatic activity. *JAK2*V617F occurs in JH2 reducing its function and therefore, this inhibition is reduced causing unconstrained constitutive JH1 kinase domain activation. JAK2 is phosphorylated without cytokine stimulation resulting in activation of STAT and ERK protein through phosphorylation with subsequent enhanced transcription.

*JAK2*V617F activates cytokine receptors including MPL, EPOR and G-CSF-R. Even though all MPNs, regardless of mutational status, have dysregulated JAK-STAT signalling, distinct STAT protein activation is associated with particular MPN subtypes (Figure 1.3A). In ET, there is upregulated activity in STAT1[45] and STAT3 but diminished STAT5 activation. In PV, by

comparison, STAT5 phosphorylation is increased and a STAT5 knockout mouse model has revealed the critical role for STAT5 in PV pathogenesis[46,47]. STAT3 and STAT5 activity is reduced in MF.

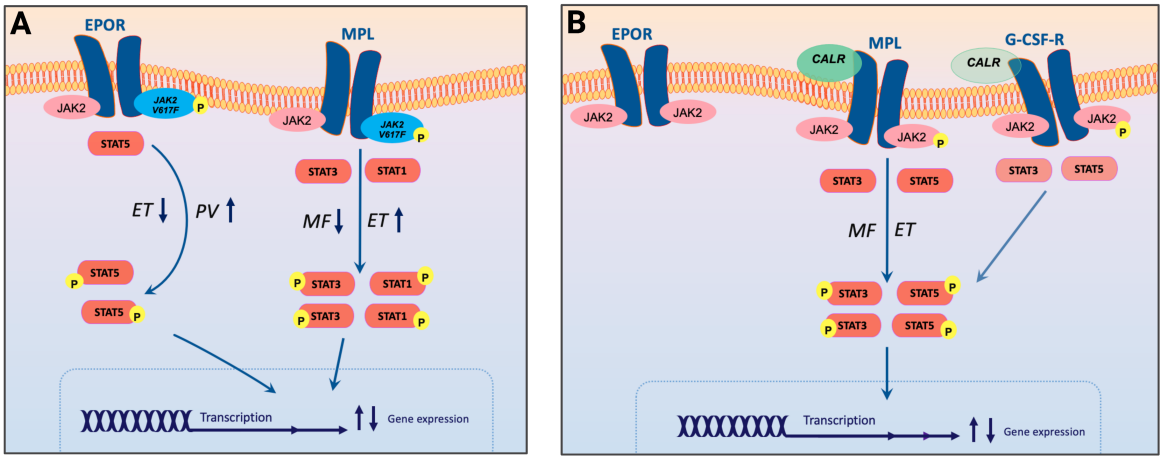


Figure 1.3. Cytokine receptor activation in MPNs. (A) Differential STAT protein activation in *JAK2*V617F MPNs. (B) In *CALR*-mutant MPN, CALR binds to MPL, signalling through JAK2 to upregulate STAT3 and STAT5 and weakly binds G-CSF-R activating STAT proteins. Adapted from O'Sullivan, J, Mead A. *Adv Biol Reg* 2019[1].

CALR mutations also act through binding cytokine receptors but predominantly activate MPL[36,37], weakly G-CSF-R and importantly, do not activate EPOR (Figure 1.3B). This receptor selectivity elucidates the *CALR* clinical phenotypes of ET and MF but never PV. *CALR*-mutant activation of MPL causes increased JAK-STAT signalling[8] through JAK2 phosphorylation. Knockdown of *JAK2* in CALR-mutant progenitors leads to a failure of megakaryocyte colonies growth[37]. *CALR* mutations activate STAT3 and STAT5 proteins[48] and also signal through MAPK pathways[49]. *MPL* mutations, *M515L* and *M515K*, have been described to enhance signalling of STAT3, STAT5, ERK and ATK pathways[14].

iv) Non-canonical effects

*JAK2*V617F has been reported to have non-canonical consequences in the nucleus including phosphorylation of histone H3 inhibiting binding to HP1alpha[50] and of PRMT5 preventing association with MEP50 reducing histone H2A and H4 methylation with altered target gene expression[51]. The influence of histone dysregulation on MPN disease biology has not been

clarified. Mutant-*CALR* cells have a reduced capacity to respond to ER stress and diminished activation of the pros-apoptotic pathway of the unfolded protein response (UPR)[52] supporting a mechanism for mutant-*CALR* cell resistance to apoptosis. Moreover, type 1 and 2 *CALR* mutations have been shown to have differential UPR activities[53]. Type 1 but not type 2 *CALR* exhibits depletion in endoplasmic reticulum calcium which leads to upregulation of the IRE1α-XBP1 UPR pathway. Targeting this pathway with an inhibitor caused cell death specifically in type 1 mutant cells.

1-1.4 Therapeutic strategies in MPNs

Current treatments for MPNs target vascular complications of thrombosis or haemorrhage through normalisation of blood parameters ("cytoreduction") and reduce splenomegaly/symptom burden. However, few pharmacological therapies reliably alter the natural disease course or halt disease progression. Haematopoietic allogeneic stem cell transplantation continues to be the only curative option for MPN management. Prognostic scoring systems have been developed in MPNs to risk stratify patients and help determine the optimal management plan.

Hydroxycarbamide (HC), a ribonucleotide reductase inhibitor, has been used as first-line cytoreductive therapy of high-risk PV/ET and hyperproliferation symptoms in MF for many years. It does not change the course of the disease and although there has been historical concern regarding leukaemogenic potential of this agent, this has not been borne out in large studies of HC-treated MPN patients. Therapy-related leukaemogenicity (reported at 8-40%) primarily applies where there has been the additional use of an alkylating agent such as busulfan[54] or pipobroman[55].

Due to concerns relating to disease transformation risk, recombinant interferon alpha, is recommended for first-line management of younger patients with PV/ET[56]. Pegylated interferon-alpha-2a[57] and a newer formulation-2b[58,59] have achieved complete molecular responses in some *JAK2*V617F-mutated PV and *CALR*-mutated ET[60] cases raising the possibility that these treatments may be disease-modifying. Interferon has been shown to directly target the stem cell compartment in MPN through increased cycling of quiescent HSCs and differentiation of progenitors in human MPN samples[61] and *JAK2*V617F HSC in murine models, ultimately leading to a MPN stem cell depletion[62] which may explain the mechanism it targets the malignant clone.

Anagrelide, a platelet targeting drug inhibiting megakaryocyte differentiation[63], is specifically used in ET, may increase the propensity to developing myelofibrosis. In the PT-1 study, anagrelide was associated with a 7% myelofibrotic risk at 5-years vs 2% in hydroxycarbamide-treated ET, p=0.01[64]. No risk was identified in ANAHYDRET study but this study had smaller numbers and included ET by WHO criteria[65]. More recently, a large ET study of 3721 patients signalled a heightened risk of myelofibrosis with anagrelide but this cohort included ET defined by an older classification which included WHO defined pre-fibrotic MF[66].

JAK inhibitors, primarily the first agent that was clinically approved, ruxolitinib (RUX), have transformed management of MPNs. This has been in MF[67] and PV[68] in terms of symptom and blood count improvement, spleen size reduction and significantly, a survival advantage in MF[69]. Unfortunately, there have been very few molecular responses or histological improvement in MF, an observation that might partly be explained by RUX's inability to target the MF HSC[70].

Treatment failure through resistance or intolerance in MPN is relatively common. Resistance or intolerance to first-line hydroxycarbamide is described in 15 – 24% of PV[71] and 20% of ET[72] cases. These patients have shorter survival and increased risk of disease progression. In patients with MF treated with RUX, the definition of treatment failure is not well-defined but includes those with primary resistance which is rare and those who lose response or are intolerant due to side effects[73]. Discontinuation rates have been used as an indirect method of inferring treatment failure, reported at 50-89% at 3 years[74]. RUX-related cytopenias are associated with disease progression and leukaemic transformation[74].

1-2 Heterogeneity in myeloproliferative neoplasms (MPNs)

Most patients with MPN carry the same mutation, *JAK2*V617F. Yet, one of the most interesting elements of MPN biology is how this mutation, often present in isolation, can produce such a broad range of clinical presentations from incidental diagnosis in an asymptomatic patient with ET or PV to an aggressive MF with a dismal prognosis. In addition, *JAK2*V617F is observed in 3% of healthy people with normal blood counts, an occurrence called clonal haematopoiesis (CH)[75,76].

*JAK2*V617F is linked to an increased risk of thrombosis, development of myeloid neoplasms and cardiovascular and overall mortality[77] in the setting of CH. A small proportion go on to develop an MPN. The probability of developing an overt MPN disease and also heterogeneity in clinical phenotypes is likely influenced by the cooperation of various factors such as host factors, mutant cell of origin, additional non-JAK-STAT driver mutations, bone marrow microenvironment, patient characteristics (age, gender, environmental exposures) and inherited predispositions[1].

1-2.1 Inherited and germline factors

First-degree relatives of individuals with MPNs have a significantly higher risk of developing an MPN due to shared genetic factors. This increased risk is estimated to be 5 to 7 times higher than in the general population[78]. This can be partially expounded by the fact that MPN patients have a disproportionately higher incidence of a common JAK2 haplotype[79] (known as 46/1 or "GGCC") in the population; 56% prevalence in MPN as compared with 24% in the population. This haplotype increases the risk of *JAK2*V617F AND *MPL*-mutated MPN but not *CALR*-mutated MPN[80].

Since identification of this haplotype, large genome-wide associations studies (GWAS) have been pivotal in uncovering the inherited genetic variations. These genetic variations are most often single nucleotide polymorphisms (SNPs) and have been reported in genes such as *MECOM*, *TERT* and *SH2B3*[81,82]. A large recent study coupled GWAS with single-cell technology[83] finding 7 new genetic risk traits in addition to those already recognised to determine that 17 risk traits imparted 20% of familial relative risk for MPN development. Interestingly, to examine whether these traits have biological significance, the genes were correlated with a target gene identifying a 15 gene “MPN signature”. Genes with capacity to alter HSC activity and self-renewal were included in the signature. Over 250,000 single healthy control HSCs were interrogated for expression of these genes. HSC gene signatures and MPN gene signature expression were similar. A particular risk variant, rs524137, is located ~12 kb downstream of *GFI1B*, a transcription factor which regulates HSC quiescence. In-vitro experiments showed that this trait caused decreased *GFI1B* expression and increase in long-term HSCs. The authors postulated that these risk traits may predispose to MPN development through expansion of the HSC pool raising the chances that a single HSC may gain a mutation.

An important observation has been recently uncovered by a number of groups that provides novel insights into MPN inception and development and furthermore, challenges the current model of cancer evolution. Mutations have been observed to occur at a steady rate, a phenomenon named "the molecular clock"[84]. Harnessing this process allowed determination of the evolutionary history of MPN HSCs, known as a phylogeny, and more significantly to track the timing of mutation acquisition in patients[85,86]. WGS was performed on single-cell-derived colonies from two key studies[85],[86]. In an analysis of 12 MPN patients, a surprising finding was that the analysis of mutation lineage trees showed that in many patients, the mutations that drive MPN development were probably acquired during fetal development or in the early stages after birth. Four patients had the *JAK2*V617F mutation as the first event identified, and the time between when they acquired the mutation and when their MPN was diagnosed ranged from 19 to 54 years[85]. The same approach was used to study two MPN patients with comparable findings that driver mutations were likely acquired early and had a protracted latency[86].

In addition, a pair of monozygotic twins who both developed MPN in their 40s were studied using WGS and single-cell analysis[87]. This work supports that the MPN in these twins had a shared clonal origin due to the transplacental transfer of mutant-*CALR* gene during fetal development. This data adds a further layer of evidence for the early developmental origins of MPN and implies that disease onset may be predestined, for example by germline risk traits or the driver mutation cell-of-origin. The finding that somatic mutations are acquired in utero or childhood without disease manifestation until many decades later, challenges the existing understanding of myeloproliferative neoplasms (MPN) disease pathogenesis. The rates of clonal proliferation varied amongst patients, suggesting that other variables such as gender, bone marrow niche and inherited traits are likely to play a role in disease onset (Figure 1.4A).

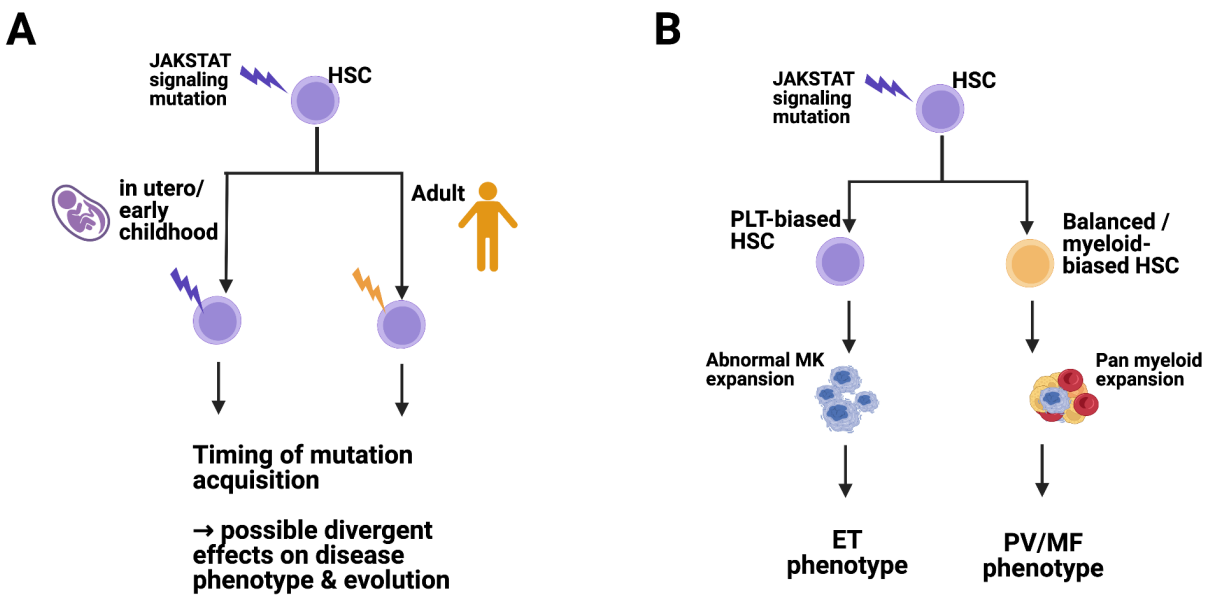


Figure 1.4. Factors influencing MPN disease heterogeneity. (A) Timing of mutation acquisition in a HSC can be as early as fetal origin to adult onset. (B) Mutation acquisition in a platelet (PLT)-biased HSC may result in an ET phenotype whereas in a balanced/myeloid-biased HSC may result in a PV / MF phenotype. Image generated using BioRender. Adapted from O'Sullivan et al. *Blood* 2022[2].

1-2.2 Target cell acquiring disease-driving mutation

Whilst mutation acquisition has been described from the apex of the haematopoietic hierarchy, the HSC, this may also be relevant to MPN heterogeneity. It is increasingly recognised that the haematopoietic compartment is heterogenous even within specific cell subtypes including the HSC. This heterogeneity includes the existence of platelet-biased HSCs which been described through murine experimental approaches: in both single-cell transplantation assays[88,89] and steady-state haematopoiesis employing barcoded lineage tracking[90]. The number of platelet-biased HSCs notably increase with age[91].

While most platelet-producing cells, megakaryocytes, develop from HSCs through intermediate multipotent progenitors, it may be possible for megakaryopoiesis to circumvent these steps and develop directly from a HSC[92]. This direct route for megakaryocyte production from platelet-biased HSCs may be used to quickly increase platelet production in emergency situations, such as acute thrombocytopenia[93].

An interesting hypothesis is that an MPN driver mutation acquired in a platelet-biased HSC might leads to an ET phenotype, while acquisition in a multipotent HSC might cause PV or MF phenotype (Figure 1.4B). However, until now, this has not been convincingly demonstrated, partly due to lack of availability of model systems to faithfully study this.

1-2.3 Clonal hierarchies and order of mutation acquisition

In genetically complex solid tumours, understanding the order in which mutations are acquired is essential to understanding cancer evolution. However, this has been historically difficult to study using bulk genomics techniques. This is more feasible in genetically simpler malignancies such as MPNs, where single-cell colony genotyping studies have shown that the order in which mutations occur can dictate disease phenotype. This concept was first highlighted by genotyping of single-cell derived colonies from MPN patients co-mutated for *JAK2*V617F and *TET2*[94], demonstrating that "JAK2-first" was associated with PV, whereas "TET2-first" was more common in ET. Similar observations in a separate study showed that "DNMT3A-first" patients had an ET phenotype[95].

A droplet-based single-cell platform, Tapestri, enables targeted mutational analysis of millions of cells simultaneously. This platform was applied to study two unusual cases where two driver mutations were detected in the same patient - *JAK2*/*CALR* and *JAK2*/*MPL* - analysing single-cell DNA of ~20,000 peripheral blood mononuclear cells[96,97]. These authors showed clearly that the mutations resided in separate clones. However, it is difficult to distinguish MPN mutation co-occurrence from a single driver mutation MPN occurring together with a 'bystander' CH.

1-2.4 The bone marrow niche

The bone marrow niche, or microenvironment, comprises the cellular and structural components that are integral to the support and maintenance of haematopoietic cells. In many hematologic malignancies, including MPNs, this microenvironment is disrupted[98]. In MPN, there is considerable cross-talk between malignant HSPCs and non-hematopoietic cells, occurring by direct cell-cell contact as well as release of pro-inflammatory/fibrotic signals, leading to a "self-reinforcing" malignant niche that promotes malignant progression[98,99]. Understanding MPN clone-niche interactions are particularly important in MF, as successful treatment may require reversal

of bone marrow fibrosis. Myofibroblast cells are thought to be directly responsible for the deposition of excess collagen and abnormal matrix proteins[100] in MF, other hematologic disorders and disease-associated fibrosis in other organs in both malignant and non-malignant pathologies[101].

Proinflammatory cytokines (e.g. TNFα, TGFβ) are elevated in MPNs, typically in MF, and are produced by both the malignant clone and stromal cells. Clonal megakaryocytes produce VEGF and TGFβ promoting angiogenesis and fibroblast proliferation, promoting development of fibrosis[102]. Elevated proinflammatory cytokines also occur in PV and ET[103] but the respective cytokine profiles only differed marginally. Preliminary data suggests the bone marrow niche may influence disease phenotype[104]. Mutated HSPCs in ET and PV occupied distinct niches, detected using combined two-photon and confocal real-time intravital microscopy. ET HSPCs moved faster and more often towards endosteal niche than PV HSPCs. Non-endosteal sinusoids were dilated in PV but not ET. These differences within the respective MPNs niches may impact MPN disease phenotypes. Though disruption of normal BM has been demonstrated in MPNs, especially in more advanced disease, it is unclear if such microenvironmental disruption is a driver of disease heterogeneity itself in patients.

1-3 Disease transformation in MPNs

1-3.1 MPN transformation presentation

While many MPN patients have indolent disease, there is an appreciable risk of disease progression to MF in the case of PV/ET and leukaemia for all MPN subtypes. ET transforming to PV has been reported, most commonly in *JAK2*-mutated disease at an incidence of 5% in a large cohort[105]. Whether the molecular basis for this is conversion from the *JAK2*V617F heterozygous to homozygous state, as has been identified in a murine model[106], remains unknown. Furthermore, there is the possibility that these cases of transformed ET represent masked PV. Disease progression to myelofibrosis after 10 years is estimated to be 4-11% for ET and marginally higher at 6-14% for PV[107]. A large study in over 2000 MPN patients, identified somatic mutations as the largest predictor of MF transformation from PV/ET[43]. Mutations in epigenetic regulators, splicing factors and RAS mutations were associated with myelofibrotic but also leukemic transformation. The authors derived a clinical prognostic model incorporating mutation

status and other relevant determinants including age to estimate the risk of disease progression for an MPN patient. In patients with ET, male gender has been associated with a higher likelihood of myelofibrotic transformation[108,3].

As mentioned, the risk of leukaemia is much higher for MF (10-20% after 10 years) than for ET and PV (1-4% after 15 years[23]). Leukaemic transformation has an especially dismal outcome with a median survival of less than 6 months[109] and limited treatment options. Traditional therapies for acute myeloid leukaemia (AML) are far less effective for leukaemic-phase MPN which may be explained by their divergent molecular landscapes[23].

1-3.2 Clonal evolution in MPN

Increasing genetic complexity correlates with progression to more advanced disease stages (myelofibrosis or leukaemia)[43]. In both PV and ET, *JAK2*V617F homozygosity increases the risk of myelofibrotic transformation[107]. Non-JAK-STAT signalling driver mutations associated with risk of progression to MF and leukaemia overlap. These include mutations in in epigenetic genes, splicing factor genes, signalling pathway genes (*NRAS, KRAS, GNAS*) and some mutations are more specific for leukaemia risk (eg. *RUNX1* and *TP53*).

The mutation spectrum in *de novo* AML differs from leukaemic-phase MPN: signalling pathway mutations are the most often mutated such as FLT3[110]. Mutations in NPM1 gene affect almost one third of de novo AML but are rarely observed in leukaemic-phase MPN[110]. Moreover, fusion events occur recurrently in de novo AML but again, are rarely present in leukaemic-phase MPN.

In myeloid neoplasms, presence of a *TP53* mutation is associated with complex cytogenetic abnormalities and dire outcomes[111,112]. In leukaemic phase MPN, *TP53* mutations are common and present in a third of cases[112,113]. However, small *TP53* clones are found in a sizeable cohort (15%) of chronic phase MPN[114] who do not transform to AML. Bi-alleleic loss of wild type *TP53* is frequent in leukaemic phase MPN[115] and other advanced myeloid neoplasms[111], suggesting a multi-hit mechanism is required for leukaemia development. Interestingly, loss of the *JAK2*V617F mutation at leukaemic transformation has been described with emergence of mutations in *TP53*.

Deciphering the clonal architecture and resolution of genetic intratumoral heterogeneity is crucial to identify therapeutic targets and effective therapies for transformed MPN. This can be achieved through single-cell methodologies. In the recent study, ~15,000 single HSPCs were analysed from 14 advanced-phase MPN cases. In all cases, there was biallelic loss of WT *TP53* in the principal clone. The 'multi-hit' mutant HSPCs contained two broad cellular subsets; one showing an erythroid gene signature, and the other enriched in leukaemia stem cell genes. The finding of expanded erythroid-biased HSPCs was not expected, as the leukaemias secondary to MPNs are not typically erythroleukaemia by standard clinical criteria[23]. WT patient HSCs showed reduction in cell cycle genes, suggestive of enhanced quiescence, and increased expression of inflammatory genes, leading to the hypothesis that a cell-extrinsic, inflammatory process mediates disease evolution and exerts a fitness advantage to the *TP53* mutant clone while repressing WT HSPCs, potentially through DNA damage-related mechanisms.

1-4 Single-cell methods and their application in MPNs

1-4.1 Single-cell genomics methods

Single-cell genomics approaches are ideally positioned to unravel the complex interplay of multiple genetic and non-genetic elements which collectively drive MPN pathogenesis. We are currently witnessing an incredibly exciting time in the field with single-cell technologies that have exponentially developed in the past decade from one of the earliest analyses in 2006 of 12 murine HSCs using microarray[116] to today, where it is possible, with extraordinary resolution to profile millions of single cells in parallel. Technologies to analyse genomic material from a single-cell also have a long history, initially through karyotype analysis – a single-cell genomics technique still in routine clinical use. Single-cell genomics have advanced at an impressive pace, from single modality analysis of a particular genomic layer – such as DNA, transcriptome/gene expression or epigenome – to the capture of multiple layers of genetic, epigenetic and also proteomic information simultaneously – in "multi-omic" techniques (Figure 1.5)[117,118]. This multi-layered information is key to unravelling the interacting elements driving cancer development, intratumoral genetic heterogeneity, evolution and resistance to therapy. These experimental advancements have necessitated parallel innovation in computational methods[119] to represent and integrate multi-modal insights at scale.

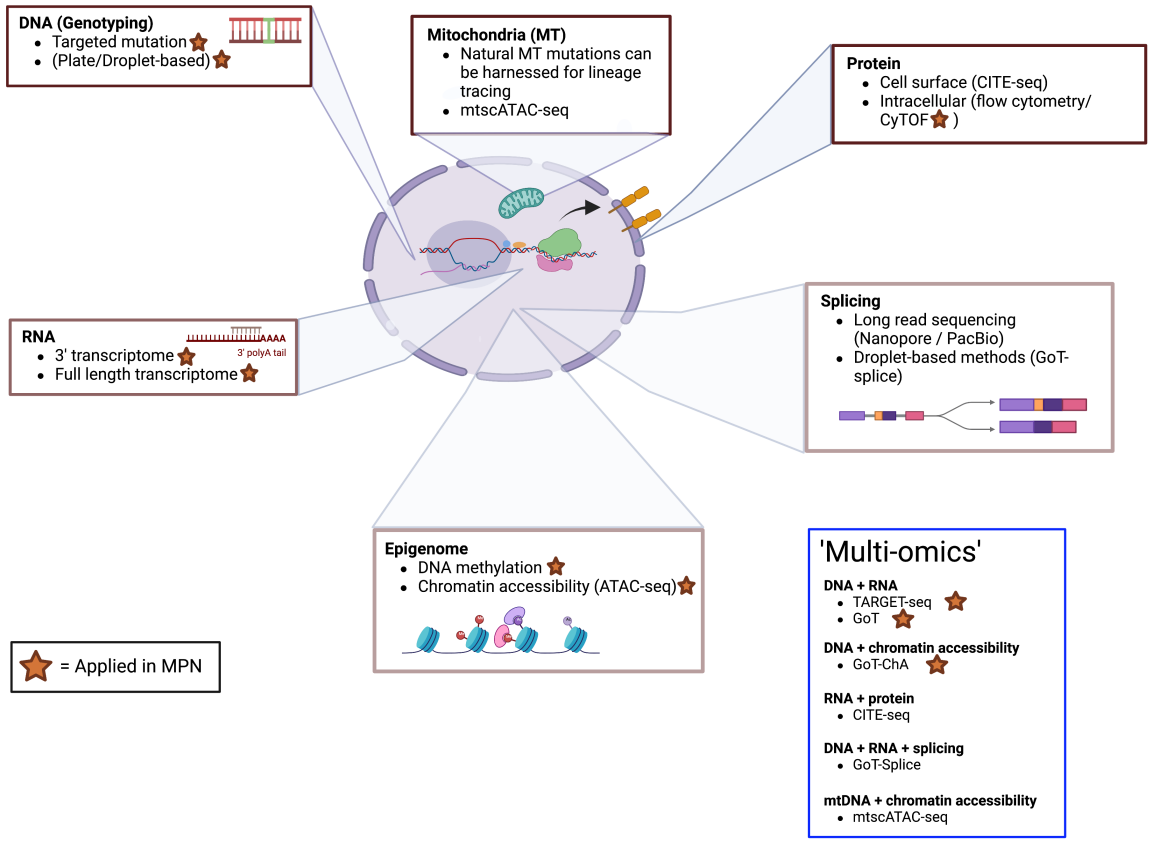


Figure 1.5. Single-cell technologies. Layers of information in the cell that can be interrogated by single-cell technologies including DNA, RNA or transcriptome, mitochondrial[120], epigenetic and splicing. Technologies combining multiple modalities, "multi-omic", are highlighted and those modalities applied to study of myeloproliferative neoplasms (MPNs) are highlighted by an orange star. Image generated using BioRender. Adapted from O'Sullivan et al. *Blood* 2022[2].

1-4.2 Single-cell genomics technologies in MPN

Simultaneous profiling of mutation status and transcriptomes is especially key in MPNs to enable interrogation of mutant-specific molecular signatures and resolve intratumoral heterogeneity. Genotyping of Transcriptomes (GoT), a droplet-based method combining these modalities, was first applied to HSPCs from patients with *CALR*-mutated ET[121]. Mutation detection was achieved with this technique by including a gene-specific primer at the first cDNA PCR step using the 10x Genomics platform. *CALR*-mutated HSPCs were found to be highly clonal, with the mutation detected throughout the HSPC compartment, the frequency highest in MkP.

However, this method could not reliably determine non-mutant cells due to allelic dropout. This has been made possible using TARGET-seq,[122,123] a highly sensitive approach enabling

simultaneous mutation and RNA-sequencing analysis in individual cells, combined with cell surface proteomics using index flow cytometry. TARGET-seq requires isolation of individual cells into 96- or 384-well plates and is lower throughput than GoT[121]. However, the high sensitivity enables reliable determination of 'true' WT cells from mutant cells where the presence of the mutation has simply not been detected due to dropouts in the sequencing reads. This method has miniaturised and automated to reduce cost, improve efficiency and throughput. This has been down through use of 384-well plates with volumes of reagents scaled down and automated liquid handling robotic has greatly increased throughput and accuracy. This precise intra-patient comparison of WT versus mutant HSPCs is a powerful approach to identify mutation-specific distinct molecular signatures, and potential therapeutic vulnerabilities (Figure 1.6)[124].

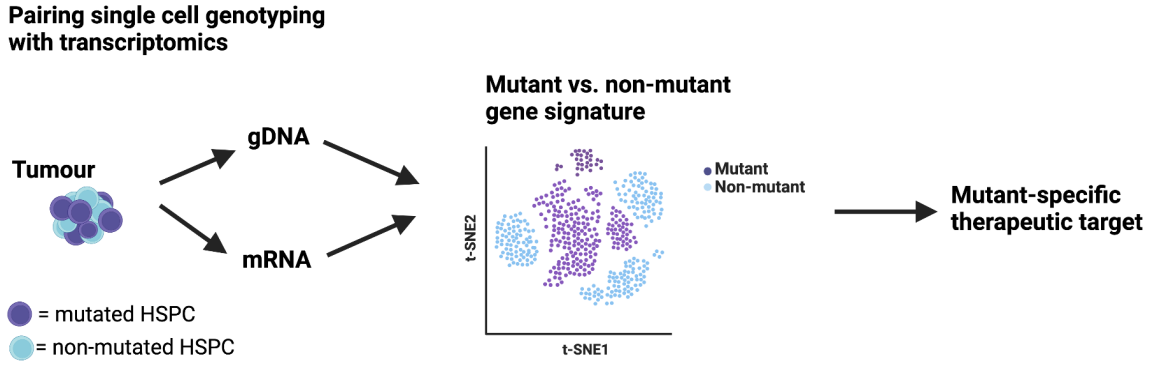


Figure 1.6. Intra-patient single-cell multi-omic analysis of mutant and non-mutant hematopoietic stem and progenitor cells (HSPC). Image generated using BioRender. Adapted from O'Sullivan et al. *Blood* 2022[2].

Application of methodologies such as TARGET-seq[122,123] can detect mutation information (gDNA) and gene expression (mRNA) to uncover mutant specific gene signatures and potential therapeutic targets.

1-4.3 New single-cell approaches

Single-cell technologies are advancing at breakneck speed, particularly muli-omic techniques. Interrogating the epigenome and spliceosome in parallel with gene expression and mutation information will be particularly powerful for dissecting the mechanisms that drive MPN evolution with improved granularity. Many of the technologies have been established using MPN to showcase their capabilities. Parallel gene expression and chromatin accessibility (ATAC-seq)[125] promises to improve our understanding of cell states and regulatory programs. GoT-ChA[126] combines mutation and epigenetic information (chromatin accessibility) in a high-throughput droplet-based platform to understand how mutant and cell-specific epigenetic regulation

influences cell states and differentiation. Studying aberrant splicing in mutant cells is also essential for identifying abnormal splicing signatures and potentially targeted therapies. GoT-Splice[127] integrates GoT[121] with long read sequencing[128] to capture full-length transcriptome and therefore combines genotyping, gene expression and splicing information of each cell. This is particularly relevant for MPNs, where mutations in splicing factors are common and resulting splice variants may be targetable.

1-4.4 Integration of single-cell genomics into clinical pathways

Although costs and technological challenges of single-cell genomics remain prohibitive, an analogy may be drawn with whole genome sequencing, which was an expensive research tool a decade ago[129] but is now offered in routine clinical care for <$1000/sample. A comparable trajectory could be expected for single-cell technologies in the coming years.

How will single-cell omics be integrated into clinical pathways? There is a clear application in improved accuracy of diagnosis, disease stage stratification, prediction and monitoring of therapy response, resistance and relapse risk. Using single-cell techniques, mutant cells can be identified and tracked during treatment, providing a more sensitive method for measuring clonal responses and clonal evolution, thereby more precisely predicting risk of relapse or disease progression. These techniques have the potential of superseding current minimal residual disease (MRD) methodologies for monitoring disease response and early detection of disease relapse. Of note, this is more readily applicable to management of acute myeloid leukaemia where MRD monitoring is routine and has direct implications for clinical decisions. Implementation in a clinical setting may be realistically possible through automation (using robotic liquid handlers) and applying higher throughput methods (such as droplet-based approaches). A next step towards this would be integration of exploratory endpoints requiring single-cell genomics in clinical trials (Figure 1.7).

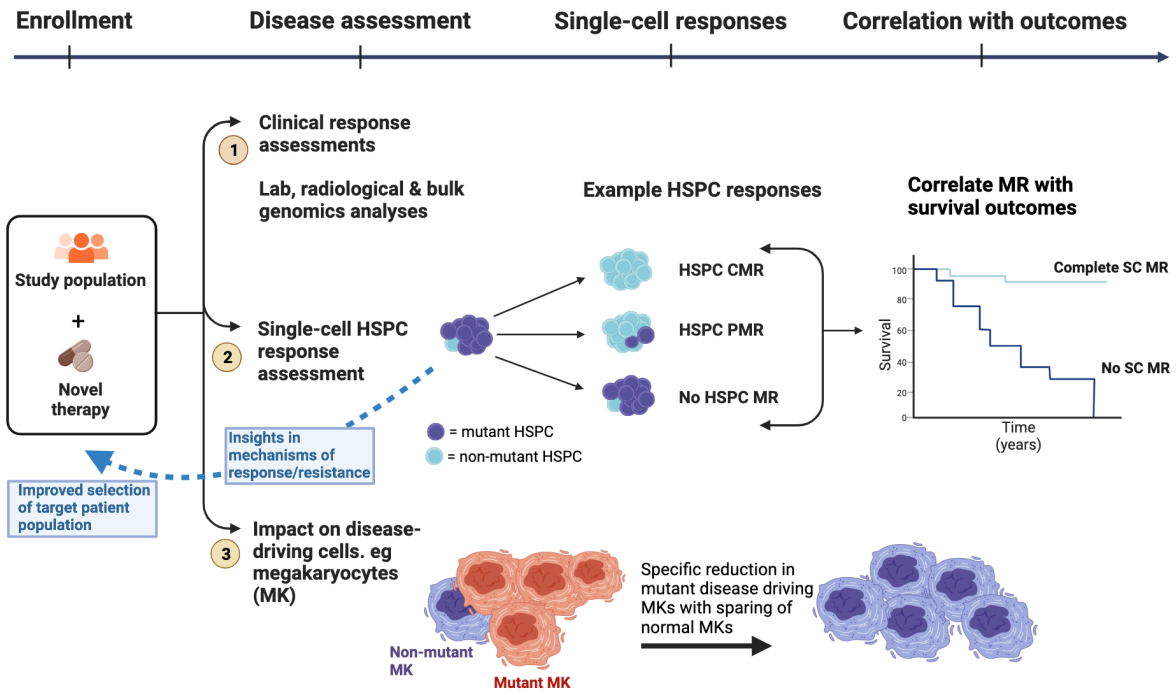


Figure 1.7. Application of single-cell (SC) technologies in clinical trials. Parallel analysis of mutant HSPCs to assess response to novel therapies and correlate with disease outcomes. HSPC CMR=complete molecular response (no mutation detected in HSPC compartment), HSPC PMR=partial molecular response (mutation reduction >50% in HSPC compartment), No HSPC MR=no molecular response (no mutation reduction in HSPC compartment). Image generated using BioRender. Adapted from O'Sullivan et al. *Blood* 2022[2].

Single-cell analyses have been conducted on clinical trial subjects in AML[130-132], but these analyses were not embedded in the trial design from study inception. Disease modification is a key aim of MPN therapies in development and ultimately requires elimination of MPN stem cells, which can only reliably be ascertained with single-cell analysis of HSPCs. This will develop an evidence-base for application in routine clinical care and realise the potential of single-cell technologies to advance personalised, genomics-driven medicine.

1-5 Hypothesis and aims

MPNs are heterogenous with various factors driving disease progression particularly somatic mutations. Currently available therapies for MPN show modest effects on modifying MPN biology or mitigating disease evolution for most patients, including the first 'targeted' drug, ruxolitinib. Despite substantial efficacy in terms of spleen and symptom responses,[133,134] improvement in bone marrow fibrosis[135] and mutation allele burdens are not common with ruxolitinib. Moreover,

treatment resistance is common and associated with poorer outcomes and disease progression. Understanding whether there are distinct mutational profiles associated with treatment resistance and furthermore, interrogating the clonal hierarchy and mutant gene signatures of the haematopoietic stem cell compartment may reveal novel therapeutic targets and most importantly, lead to development of new treatments for these patients.

Hypothesis

i) The mutational profile in treatment resistant/intolerant (RES/INT) high-risk MPN patients is distinct

ii) Disease progression is increased within distinct genetic subgroups of treatment RES/INT high-risk MPN

iii) Specific molecular pathways are dysregulated within HSPCs of distinct genetic high-risk MPN subgroups

My specific aims with this project were to:

1. Characterise the mutational profile in treatment resistant/intolerant high-risk MPN and correlate this with disease transformation risk

2. Delineate the clonal composition and haematopoietic cellular landscape of a molecularly-defined high-risk MPN subgroup

3. Interrogate the molecular signatures promoting transformation events within a molecularly-defined MPN subgroup using single-cell multi-omic methodologies

Chapter 2: Methods

2-1 Patient banking and sample selection

2-1.1 High-risk ET patients enrolled onto MAJIC-ET for mutation profiling

Inclusion criteria for the MAJIC-ET study and outcome measures were previously reported[136,137]. Briefly, 110 patients were enrolled onto the ET arm of the MAJIC study which was a basket trial design also enrolling a separate PV cohort. Patients included had high-risk ET and were resistant or intolerant to hydroxycarbamide. Well-annotated clinical information was contemporaneously collected in parallel with peripheral blood (PB) and bone marrow (BM) samples banked from all patients at baseline and serially; PB samples every 3-4 months and BM samples annually during the study period. Genomic DNA (gDNA) was isolated using DNeasy Blood and Tissue Kit (Qiagen) as per manufacturer's instructions from either PB whole blood or stored PB granulocyte pellets.

2-1.2 High-risk MF patients enrolled onto PAC203 study

Mutational analysis was performed on 110 patients enrolled onto the PAC203 study at baseline (of total 164 recruited and 161 treated), and 42 patients who had stored DNA at 24 weeks follow-up.

2-1.3 Patients within distinct genetic MPN subgroup for single-cell analyses

The cohort or patients for single-cell analysis was focused on a specific mutated-subgroup and the selected group extended beyond MAJIC-ET to achieve sufficient patient numbers for biologically meaningful comparisons. This comprised of patients co-mutated with *JAK2*V617F and a second specific mutation, patients with *JAK2*V617F alone and age- and tissue-matched healthy controls. Patients were matched for age, treatment and disease subtype.

2-2 Bulk mutation analysis

Next generation sequencing (NGS) was performed on PB genomic DNA (gDNA) in all patients on the MAJIC-ET and PAC203 cohorts. In MAJIC-ET, these were done at baseline (n=110) and in patients with samples available at a later time point. Driver mutations were sequenced using a

NGS assay as previously described[138]. This sequencing was performed by the molecular diagnostics laboratory at Oxford University Hospitals. Mutational profiles of NDM were analysed using an International Organization for Standardization Standardization (ISO 15189:2012) accredited Illumina TruSeq Custom Amplicon Panel including 32 gene mutation hotspots & exons frequently mutated in myeloid malignancies (~56,000 bp, 341 amplicons); *ASXL1, ATRX, DNMT3A, EZH2, TET2, CEBPA, ETV6, NPM1, PHF6, RUNX1, SETBP1, SF3B1, SRSF2, TP53, U2AF1, WT1, ZRSR2, CBL, CBLB, CBLC, CSF3R, FLT3, HRAS, JAK2, KIT, KRAS, MPL, NRAS, PDGFRA, PTEN, IDH1, IDH2*[139]. Paired-end indexed libraries were prepared for each sample, sequenced on the Illumina MiSeq/NextSeq platforms and FASTQ files were aligned to the reference genome (GRCh37/hg19). The minimum depth was ≥100 reads per base and minimum accepted coverage was achievement of this depth in ≥95% of targeted bases. Alignment and variant calling were performed in Basespace (utilizing BWA and GATK/Somatic variant caller; Illumina), while filtering and annotation were performed using a combination of Variant Studio (Illumina) and a custom designed in-house algorithm to evaluate the variants identified.

Pathogenicity of each variant was determined using Exome Aggregation Consortium (ExAC) population frequencies, the Single Nucleotide Polymorphism database (dbSNP), the Catalogue Of Somatic Mutations In Cancer (COSMIC) databases and published literature. Variants considered pathogenic included if reported as somatic in COSMIC and not found in germline databases, present at the same genomic location of a somatic mutation reported in COSMIC, not reported in dbSNP databases occurring within 3 amino acids of a reported variant in COSMIC with a variant allele frequency (VAF) <40-45% and predicted to result in a truncated protein. Single-nucleotide variants (SNVs) were excluded if they had a population frequency of >1% or had a population frequency of <1% but with ethnicity bias and a variant allele frequency (VAF) close to 50%. Variants determined to be pathogenic were reported if present with a VAF of ≥5%.

2-3 Cytokine assay in a high-risk MF cohort

Cytokines were analysed on using the Myriad RBM platform, a microsphere-based immuno-multiplexing technology; B2M, BDNF, CRP, ETOAXIN1, FABPHE, FACTOR7, GH, GMCSFA, ICAM1, IFN- γ, IL-10, IL-12P40, IL-12P70, IL-15, IL-17, IL-18, IL-1ALP, IL-1-β, IL-1RA, IL-2, IL-23, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, insulin, leptin, LOX1, LPA, MCP1, MIP1AL, MIP1BE, MMP3,

MMP9, NTPROBN, SAP, SCF, TBG, TM, TNFα, TNFβ, TSH, VEGF, vitronectin and VWF. This assay was performed by an external partner (CTI BioPharma).

2-4 Antibody-derived tag (ADT) selection for CITE-seq

For CITE-seq[140], ADTs were selected to delineate cell populations in high-throughput scRNA-seq (10X genomics). Commercial TotalSeq-A ADTs (Biolegend) were used which have a polyA sequence that can bind to polydT capture methods (Figure 2.1). ADTs were selected that would identify and improve resolution in distinguishing HSPCs and mature cell populations (lymphoid and myeloid) beyond that achievable by RNA-sequencing (Table 2.1). Immune checkpoint markers and major histocompatibility complexes (MHC) class I and II were included to additionally explore immune dysregulation in MPN (Table 2.1). Each ADT was included at 30% of Biolegend's recommended concentration; 0.3μg (0.6μL) added to 100μL staining volume. This was based on data establishing optimal titration of 52 ADTs[141] finding that most ADTs reached a saturation plateau at 0.062 - 0.25μg per 100μL. At concentrations above this, the background noise increased and reduced signal-to-noise ratio.

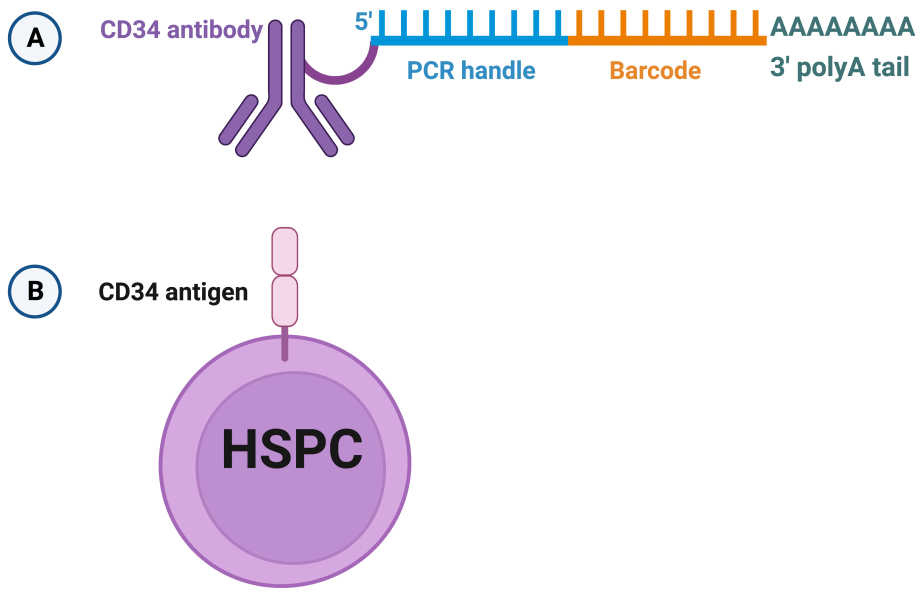


Figure 2.1 Schema of ADT construct. The antibody is conjugated to an oligonucleotide sequence which starts with a PCR handle (blue) at the 5' end, antibody-specific barcode sequence (orange) and polyA tail at 3' end (turquoise) (A). CD34 antibody is illustrated which recognises and binds to CD34 antigen expressed on haematopoietic stem and progenitor cells (HSPC). Image generated using BioRender.

Antibody	Target cell type/class	Clone
CD34	HSPC	581
CD117 (c-kit)	HSPC	104D2
CD133	HSPC	clone 7
CD164	HSPC	67D2
CD38	HSPC	HB-7
CD45RA	HSPC	HI100
CD90 (Thy1)	HSPC	5E10
CD123	HSPC/Dendritic	6H6
CD41	MkP	HIP8
CD42b	MkP	HIP1
CD61	MkP	VI-PL2
CD62P (P-Selectin)	MkP	AK4
CD151 (PETA-3)	MkP	50-6
CD36	ErP/MkP	5-271
CD71	ErP	CY1G4
CD235ab	ErP	HIR2
CD66b	Myeloid	6/40c
CD25	Dendritic	BC96
CD83	Dendritic	HB15e
CD14	Monocytic	63D3
CD16	Monocytic	3G8
CD47	Macrophage	CC2C6
CD172a (SIRPα)	Macrophage	15-414
CD56 (NCAM)	NK	5.1H11
CD57	NK	QA17A04
CD2	T cell	TS1/8
CD7	T cell	CD7-6B7
CD4	T cell	RPA-T4
CD8a	T cell	RPA-T8
CD84	T cell	CD84.1.21
CD69	T cell	FN50
CD70	T cell	113-16
CD10	B cell	HI10a
CD20	B cell	2H7
CD193 (CCR3)	Eosinophil/Basophil/T cell	5E8
CD152 (CTLA4)	Immune checkpoint	BNI3
CD273 (B7-DC, PD-L2)	Immune checkpoint	24F.10C12
CD274 (B7-H1, PD-L1)	Immune checkpoint	29E.2A3
CD223 (LAG-3)	Immune checkpoint	11C3C65
CD366 (Tim-3)	Immune checkpoint	F38-2E2
CD184 (CXCR4)	Immune checkpoint	12G5
HLA-A,B,C	Immune system	W6/32
HLA-A2	Immune system	BB7.2
HLA-DR, DP, DQ	Immune system	Tü39
LAP (TGF-β1)	Cytokine	TW4-2F8
CD126 (IL-6Ra)	Cytokine receptor	UV4

Table 2.1. Totalseq cite-seq antibody markers included.

ADTs were combined in a pool prior to subsequent cell staining. Each ADT was vortexed briefly and spun down to reduce ADT aggregates and 0.6μL per ADT per sample was combined into a pool in a 1.5mL Eppendorf tube. Isotype controls were included. Each antibody pool was topped up with 10X-specific FACS buffer (5% FCS, 2mM EDTA in PBS) to a total volume 50μL. Each 50μL ADT pool was spun at 14,000g for 10 minutes at 4C prior to adding to later adding to cells suspended in 50μL to a final total staining volume of 100μL.

HTO (Totalseq-A, Biolegend) were also included in the experimental design. These have the same construct as ADT markers except these are antibodies which target antigens which are ubiquitously expressed on all cells so that they can be used as sample barcodes. One HTO was used per sample and each HTO barcode could be detected through sequencing and used for sample demultiplexing.

2-5 Fluorescence-activated cell sorting (FACS) and single-cell isolation of HSPCs

2-5.1 Sample preparation and staining

HSPCs are rare cells in the bone marrow (BM) accounting for ~1% of BM cells in healthy individuals. These are rarely found in the circulation but in MPNs, predominantly MF, these circulate at an increased frequency. Taking advantage of this, cryopreserved samples from PB or BM were processed for sorting depending on sample availability.

Firstly, sample cryovials (2-3 at a time) were thawed quickly in a waterbath at 37C. One mL of FCS was added dropwise to each sample and transferred to a 15mL Falcon tube. Each sample vial was washed with 1mL of thaw media (IMEM with 20% FCS and 10% DNaseI added) and transferred to the Falcon tube which was finally topped up with thaw media to 15mL. The Falcon tube was centrifuged at 350g for 10 minutes and the supernatant discarded. The pellet was carefully resuspended in 5mL of FACS buffer for one additional wash step and the cells were counted using an haematocytometer.

An aliquot of cells from each sample (patients and controls) was taken for fluorescent minus one controls (FMOs), mixed together in a new 1.5mL Eppendorf tube and then divided to obtain 100-150,000 cells per 1.5mL FMO tube. FMOs were included in each experiment to set gates for each

fluorophore. Single-stained controls were run for the first experiment only on each instrument using compensation beads (CompBeads, BD Biosciences) to set laser voltages and establish the spectral overlap for each fluorophore to set appropriate compensation. FACS was performed using a BD Fusion II (Becton Dickinson) and SONY SH800S instruments for bulk enrichment and 384-well plate sorting respectively (Figure 2.1).

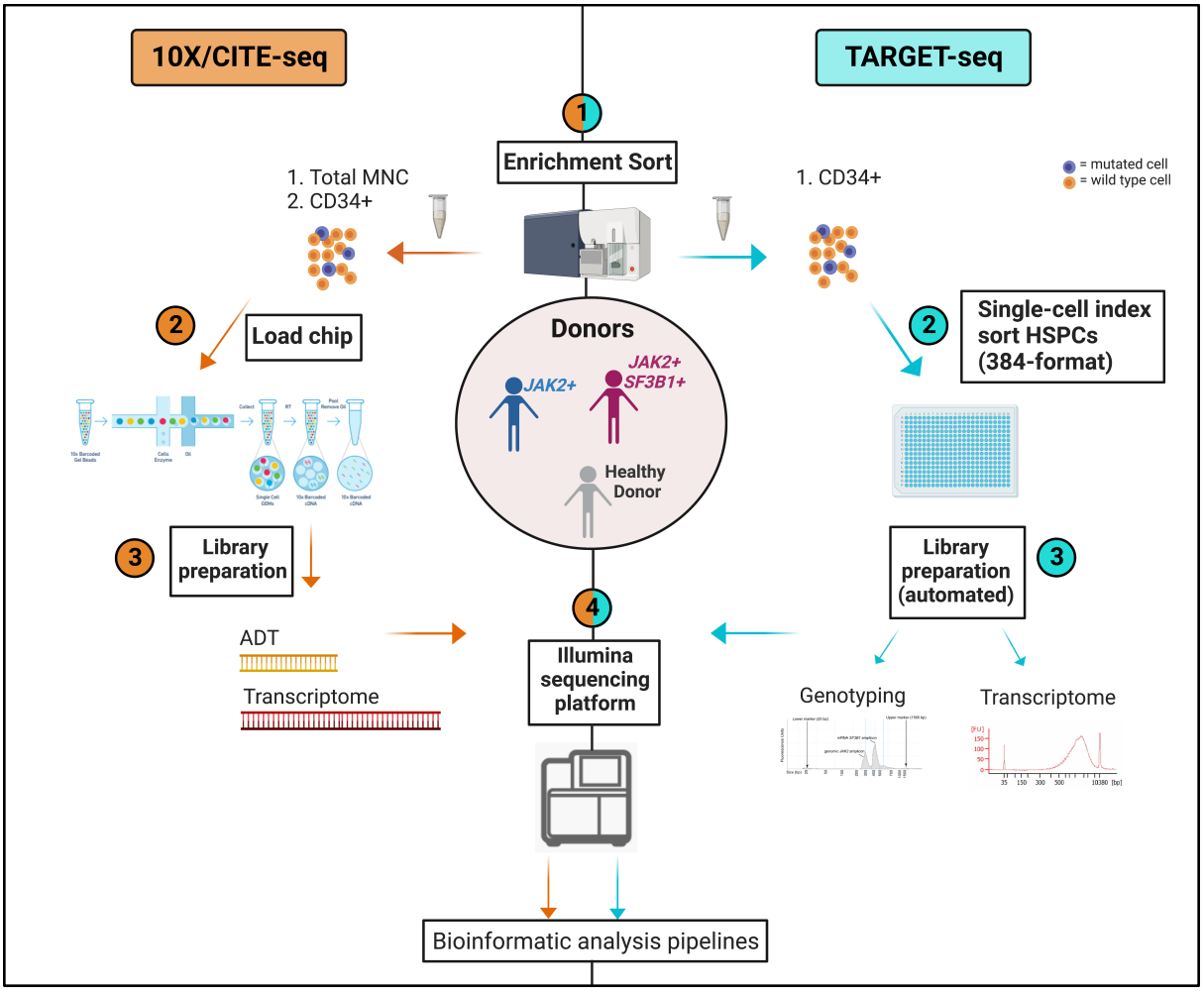


Figure 2.2. Schema for 10X genomics/CITE-seq and TARGET-seq sorting, library preparation and sequencing. 1. Donor samples are separately enriched for 10X/CITE-seq (orange) sorting total MNC and CD34+ cells and in parallel CD34+ cells are sorted for TARGET-seq (turquoise). 2. The cells are loaded onto 10X chromium chip (orange) and for TARGET-seq cells are further sorted as single-cells into 384-well plate (turquoise). 3. Library preparation to generate ADT and transcriptome libraries for 10X genomics/CITE-seq (orange) and genotyping and transcriptome libraries for TARGET-seq (blue). 4. Sequencing performed on Illumina platforms. Image generated using BioRender.

Droplet-based 10X genomics/CITE-seq and plate-based TARGET-seq were performed on each sample. Therefore, at this stage prior to antibody staining, the samples were divided with 5 million cells taken for 10X genomics/CITE-seq and the remaining cells for TARGET-seq. Each sample was then centrifuged at 350g for 10 minutes and the supernatant discarded.

For 10X genomics/CITE-seq, each sample was resuspended in 50μL 10X-specific FACS buffer. Human Fc-block (1:200; BD, Cat# 564220) was added and incubated with the samples for 5 minutes at 4C. Staining was performed by adding the following sequentially to aim a final staining volume 100μL; CD34+ APC-efluro780 fluorophore (1μL), CITE-seq antibody pool (50μL) and HTO antibody(1μL). The samples were incubated for 30 minutes in the dark at 4C following by two wash steps by adding 1mL of 10X-specific FACS buffer, centrifuging at 400g for 5 minutes and resuspending to a final volume of 400μL.

Antibody	Flurophore	Company	Dilution	Clone
CD34	APC-e780	eBiosciences	1:150	4H11
CD38	PETxR	Invitrogen	1:21	HIT2
CD123	PE-Cy7	eBiosciences	1:60	6H6
CD45RA	PE	BioLegend	1:150	HI100
CD90	BV421	BioLegend, IgG1	1:30	5E10
CD41	APC	eBiosciences	1:37.5	HIP8
CD71	AF700	BD Pharmigen	1:30	M-A712
CD8 (Lineage)	FITC	BioLegend	1:100	RPA-T8
CD10 (Lineage)	FITC	BioLegend	1:30	HI10a
CD20 (Lineage)	FITC	BioLegend	1:150	2H7
CD66b (Lineage)	FITC	BioLegend	1:15	G10F5
CD127 (Lineage)	FITC	eBioscience	1:30	eBioRDR5
CD2 (Lineage)*	FITC	eBioscience	1:15	RPA-2.10
CD3 (Lineage)*	FITC	eBioscience	1:15	OKT3
CD14 (Lineage)*	FITC	eBioscience	1:15	61D2
CD16 (Lineage)*	FITC	eBioscience	1:15	CB16
CD19 (Lineage)*	FITC	eBioscience	1:15	HIB19
CD56 (Lineage)*	FITC	eBioscience	1:15	TULY56
CD235a (Lineage)*	FITC	eBioscience	1:15	HIR2

Table 2.2. HSPC immunophenotype panel used for FACS. *pre-mixed pool of antibodies available as "Human hematopoietic lineage cocktail" (Cat **#22-7778-72).**

For TARGET-seq, each sample was resuspended in 90μL (< 10 million cells) or 180μL (≥ 10 million cells) of FACS buffer in a 1.5mL Eppendorf tube. Again, Human Fc-block (1:200; BD, Cat# 564220) was added and incubated with the samples for 5 minutes at 4C. The HSPC antibody panel for TARGET-seq sorting (Table 2.2) was made up as a pool and 59uL of this pool was added to each sample to a volume of 150μL. In parallel, cells for each FMO was stained with the relevant antibody mix to a volume of 75μL. All samples were stained for 20 minutes at 4C and then washed once with 1mL of FACS buffer (IMDM with 10% FCS), centrifuged at 400g for 5 minutes and resuspended in 400μL of FACS buffer. The stain for dead cell exclusion, 7-AAD, was added to both TARGET-seq and 10X/CITE-seq samples at the sorter at a concentration of 1:100.

2-5.2 10X genomics/CITE-seq FACS

MNCs and total CD34+ cells were sorted separately for 10X genomics/CITE-seq into PCR grade microfuge 1.5μL containing 2μL non-acetylated 10% BSA (Cell Stem Tech Catalogue #09300 in PBS, filtered twice (0.2μm)) (Figure 2.3).

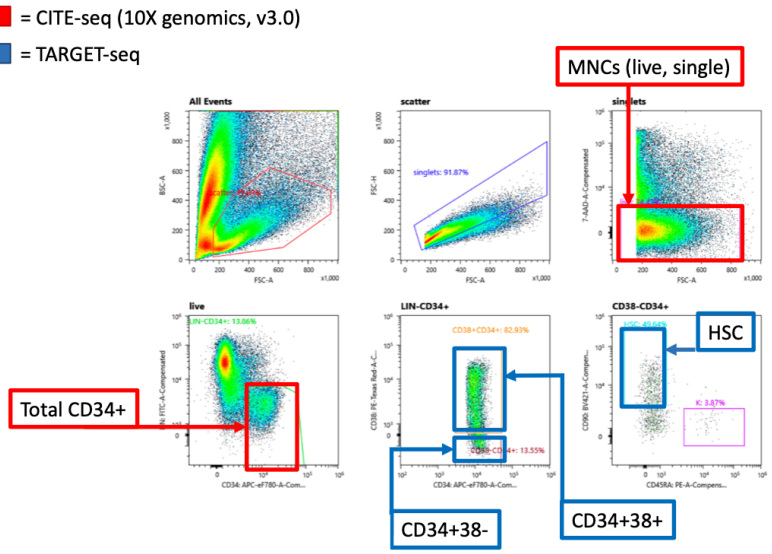


Figure 2.3. Populations sorted at enrichment. Total CD34+ and MNCs were sorted for 10X genomics, total CD34+ for TARGET-seq enrichment. For single-cell 384-well plate TARGET-seq sorting, CD34+CD38-, CD34+CD38- cells and HSCs were sorted.

To mitigate against batch effect, donor MNC and CD34+ populations were mixed after sorting (Figure 2.4A) and loaded onto a 10X Chromium Next GEM Chip G (Figure 2.4B). Demultiplexing was performed computationally downstream after sequencing.

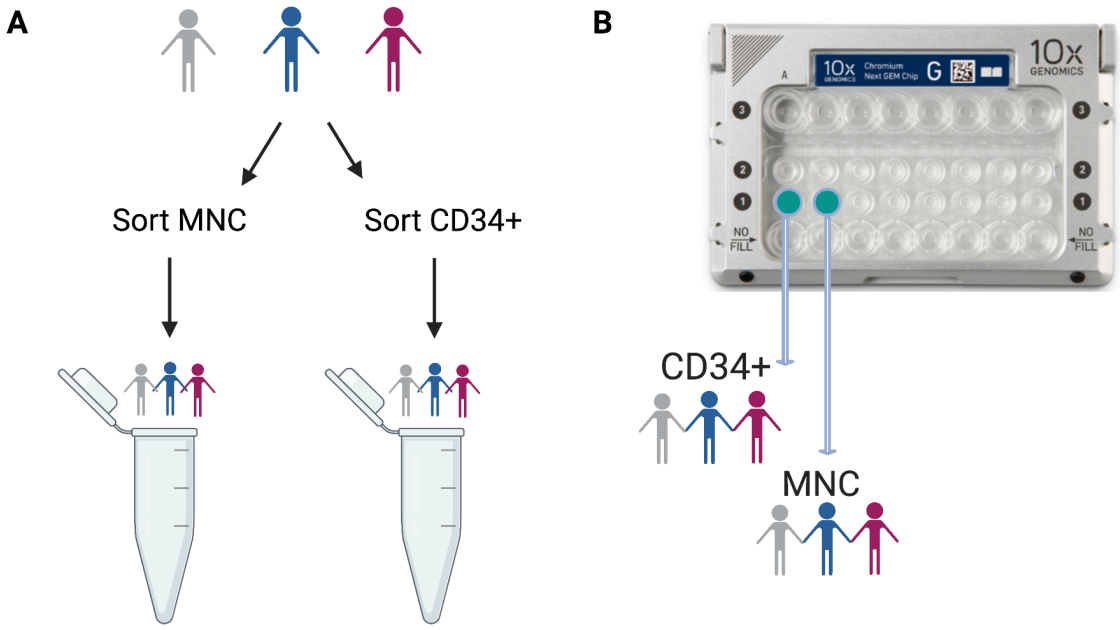


Figure 2.4. Donor mixing strategy for 10X genomics/CITE-seq. (A) Three donors combined for each population sorted into a PCR tube and (B) then loaded into a lane on 10X Chromium Next GEM Chip G. Image generated using BioRender.

2-5.3 TARGET-seq FACS

Samples underwent lineage-CD34+ bulk enrichment sort on BD Fusion II sorter into Eppendorf tubes prior to single-cell 384-well sorting. This was performed to reduce the sort time per 384-well plate; longer than 20 minutes per plate increases risk of RNA degradation.

For single-cell 384-well plate sorting, accurate alignment was required to ensure a single cell was sorted into the bottom of each well. To achieve this consistently, I chose a custom sorting purity mode in which 50% of the previous and following droplet to the target one is empty, and the target event is within the 75% central part of the target droplet[123] (Figure 2.5A). This maximised cell recovery whilst making sure only one cell per well was sorted. To set up the alignment, a test sort was first performed using a 384-well plate covered with a PCR film sorting 50 droplets into the corners of the plate. The alignment was adjusted until the drops fall into the centre of the wells. Once correct, this step was repeated without a PCR film sorting 20 droplets into the plate corners. The alignment was correct once the drops were positioned into the bottom of each well with no spray of liquid on the side of the well.

Single HSPCs were sorted into 192 wells of the plate, again, to minimise the sort time (Figure 2.5C) and index sorted to record the fluorescence intensity of each marker included in the HSPC panel for additional cell surface proteome information. A mixture of lineage-CD34+, CD34+CD38-, CD34+CD38- progenitors and HSCs were sorted; the proportions sorted were dictated by the frequency the populations per sample.

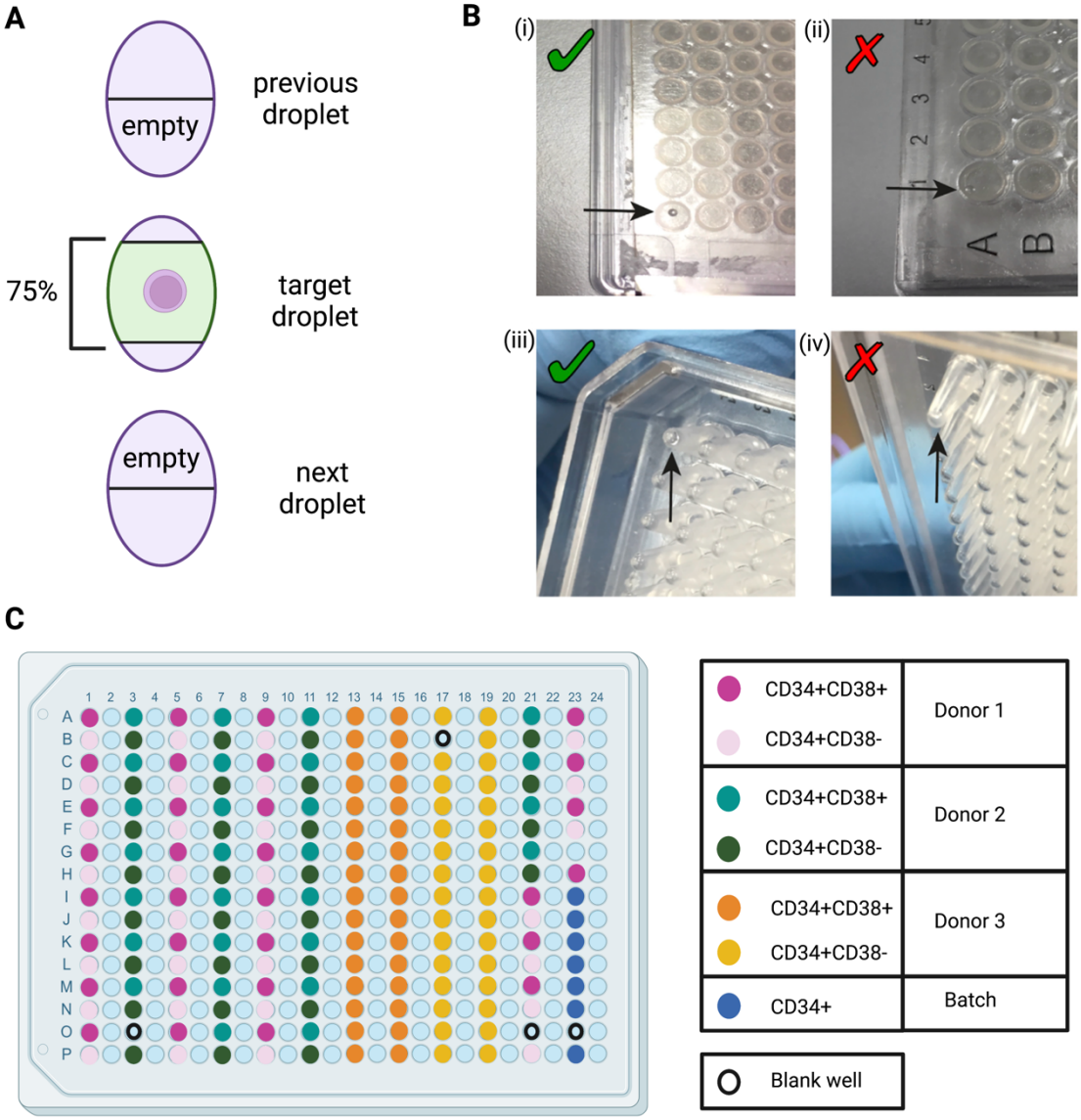


Figure 2.5. 384-well single-cell sorting. (A) Sony custom single-cell purity settings. (B) (i) Drop correctly aligned on center of film-covered well, (ii) Drop incorrectly aligned at the edge of film-covered well. (iii) Drop correctly aligned at the center of the bottom of the well (iv) Drop incorrectly aligned to the side of the well (adapted from STAR protocols[123]). (C) Example layout of sorted cells into 384-well plate with cells sorted coloured by type and donor. Image generated using BioRender.

To mitigate batch effect caused by sort and plates, patient and healthy control samples were sorted onto each plate (Figure 2.5C) and cells from a separate healthy donor was sorted into a column of each plate for every experiment. Blank wells were included (n=4 per plate) to set the threshold for genotyping background "noise" (Figure 2.5C). Each well of the plate contained 2μL of TARGET-seq lysis buffer. Each plate was covered with an aluminium film after sorting, briefly vortexed, spun down, snap frozen on dry ice and stored at -80C until the next step. The sorting experiments were performed with the assistance of Dr. Sally-Ann Clark at the Flow Cytometry Facility in the Weatherall Institute of Molecular Medicine.

2-6 Single-cell methodology I: CITE-seq/10X Chromium (droplet-based)

Single-cell 10X/CITE-seq libraries were prepared by Dr. Neil Ashley in the Single-Cell Facility in the MRC Weatherall Institute of Molecular Medicine using Chromium Next GEM Single Cell 3' Reagent Kit v3.1 (user guide CG000206_Rev_D). After loading the chip wells, exomic libraries were prepared using the published 10X genomics protocol for cDNA generation but the method was modified to prepare CITE-seq libraries with Totalseq-A antibodies (Biolegend). At the cDNA amplification step, ADT and HTO additive primers were included (sequences in **Appendix 1**) in the mix at 1μL (0.2μM) each and 35μL of post RT sample was added to this with the protocol followed for cDNA amplification in user guide CG000206_Rev_D. At the cDNA clean up step, the cDNA and SPRIselect beads were mixed, incubated and then placed on the magnet, the supernatant that is usually discarded is kept. This fraction contains the ADT and HTO libraries. For ADT and HTO amplification, two separate reactions were performed taking 5μL of the fraction for each master mix (Table 2.3 (i-ii). The ADT and HTO reaction master mixes were incubated on a thermocycler (Table 2.3 (iii-iv)).

Table 2.3 (i) ADT PCR reaction

Mastermix	**Volume (μL)**
ADT fraction	5
SI PCR primer (10uM)	2.5
TrueSeq Small RNA RPIx (10uM)	2.5
2X Kapa Hifi Master Mix	50
RNAse-free water	40
TOTAL	**100**

Table 2.3 (ii) ADT PCR reaction

HTO PCR	**Volume (μL)**
HTO fraction	5
SI PCR primer (10uM)	2.5
TrueSeq D70x_LONG RPIx (10uM)	2.5
2X Kapa Hifi Master Mix	50
RNAse-free water	40
TOTAL	**100**

Table 2.3 (iii) ADT		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
98	02:00	1
98	00:20	15
60	00:30	
72	00:20	
72	05:00	1
4	HOLD	HOLD

Table 2.3 (iv) HTO		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
98	02:00	1
98	00:20	13
64	00:30	
72	00:20	
72	05:00	1
4	HOLD	HOLD

Table 2.3. ADT and HTO PCR reactions. (i) ADT PCR reaction mastermix. (ii) HTO PCR reaction mastermix. (iii) ADT PCR thermocycler conditions. (iv) HTO PCR thermocycler conditions.

ADT/HTO clean-up was performed as per manufacturer instructions at a ratio of 1.2:1 beads to DNA adding 120μL SPRI select (Beckman Coulter, Cat# B23317), incubating for 5 minutes and performing two 80% ethanol washes. The cleaned libraries were eluted in 40.5μL elution buffer (EB). Exomic (cDNA), ADT and HTO library qualities were checked using Agilent Bioanalyzer; High Sensitivity DNA Kit (Agilent, Cat# 5067-4626). Exomic libraries had expected peak 400-600bp (Figure 2.6A). ADT and HTO libraries were expected to have a peak at 180 bp (Figure 2.6B-C). Some ADT/HTO libraries had longer bp peaks at 400-500bp. (Figure 2.6D). This was caused by overamplification of the library which leads to depletion of primer availability causing self-priming PCR products (double ADT or HTO sequences) called “bubble products”. These can be sequenced but can interfere with library quantification.

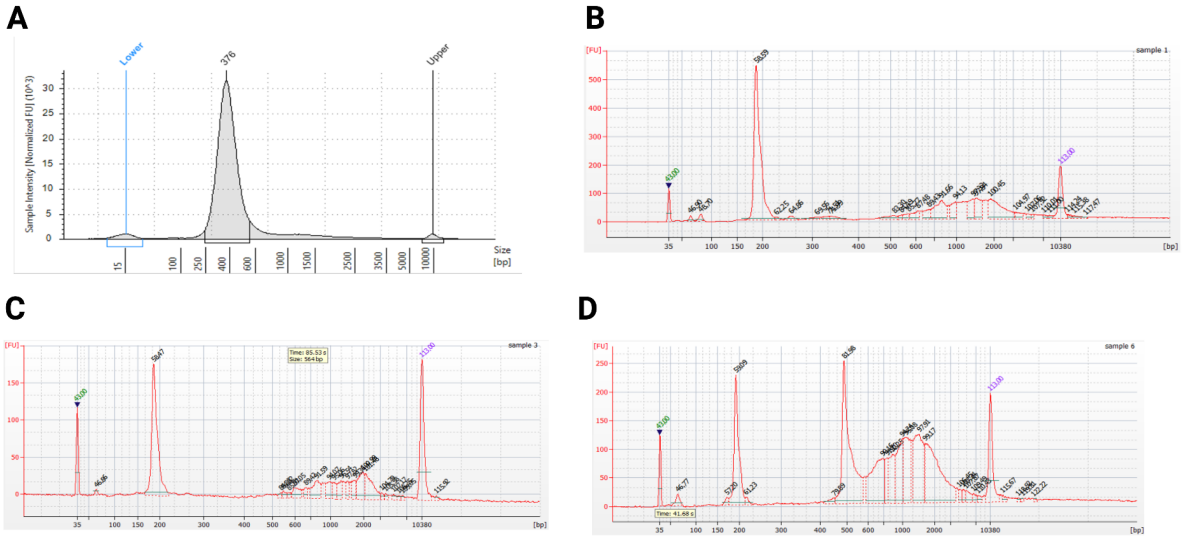


Figure 2.6. Exomic, ADT and HTO library quality traces. (A) Exomic library example (Bioanalyser). (B) ADT library example. (C) HTO library example. (D) Bubble product trace with second large peak at ~450bp.

Therefore, an additional 2 cycles ADT/HTO PCR were performed with P5/P7 generic primers to enrich for ADT/HTO libraries. Libraries were quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Cat# Q32854). Exomic, ADT and HTO libraries were compatible with Illumina sequencing primers and were sequenced at 50,000, 5000 and 1000 reads per cell respectively on a HiSeq / Novaseq Illumina platform. The libraries were sequenced together with a sequencing configuration: 28 cycles for read 1, 8 cycles for the index read and 91 cycles for read 2.

2-7 Single-cell methodology II: TARGET-seq (plate-based)

2-7.1 Lysis buffer preparation

Lysis buffer was prepared on the same day in a pre-PCR area to reduce risk of contamination as cell sorting. Moreover, to improve consistency and avoid cross-contamination, this step was automated on a MANTIS Liquid Handling Platform (FORMULATRIX) and a VIAFLO 96/384 (INTEGRA Biosciences)[123]. The lysis buffer master mix was prepared to make up 10 "half" (192 well) 384-well plates (Table 2.4). The lysis buffer master mix was aliquoted into 192 wells of a 384-well plate using the MANTIS; 17μL of liquid was dispensed using a MANTIS high volume (HV) chip (FORMULATRIX, Cat# MCHVSMR6). Next, 2μL of lysis buffer master mix from each well of the stock 384-well plate was dispensed into each well of 10 half 384-well plates using the VIAFLO 96/384 INTEGRA electronic pipette (INTEGRA Biosciences). Each plate was centrifuged after aliquoting, covered with a clear PCR film and kept on ice until used for single-cell sorting.

Lysis buffer master mix	1 Cell (Volume μL)	1,920 Cells + 13% Dead Volume (μL)
Triton 0.43%	0.88	1917.5
RNAase Inhibitor	0.05	109
dNTPs (10mM)	0.5	1090
Protease	0.05	109
ERCC E6	0.02	43.2
TOTAL	**1.5**	**3269**

Table 2.4. Lysis buffer mastermix. Mastermix prepared for 1920 cells sufficient for 10 plates (192 cells per plate) with 13% dead volume.

2-7.2 Single-cell cDNA synthesis

In this step, mRNA is retrotranscribed to cDNA. This is a critical step as mRNA can degrade during the RT step due to time delay, handling or contamination. This was performed in a pre-PCR clean area which is free from amplified PCR material to reduce contamination. Until starting this step, plates were kept on dry ice. Heat inactivation is the first step to denature the protease enzyme preventing it inhibiting any subsequent steps. The protease is included to release genomic DNA from histones for single-cell genotyping. Prior to proceeding, non-enzyme components of RT were added to a RT buffer master mix; water and mRNA primers (Table 2.5).

RT mastermix	**1 cell (μL)**	**222 cells + 15% Dead Volume (μL)**
Buffer 5X	1	222
RNAse Inhibitor*	0.125	27.75
TSO-LNA*	0.05	11.1
mRNA primers (F+R) 200uM	0.035	7.77
SMARTScribe Enzyme*	0.5	111
Water	1.29	286.38
TOTAL	**3**	**666**

Table 2.5. RT mastermix. * = reagents added during heat inactivation stepin the order listed. F=forward primer, R=reverse primer.

Heat inactivation was then performed for 15 minutes at 72ºC in a PCR thermocycler. During heat inactivation, RNase inhibitor, TSO-LNA (100μM) and SMARTScribe (RT enzyme) (Table 2.5) were added to the RT buffer master mix. Once completed, the master mix was dispensed into each plate using the MANTIS and then immediately each plate was incubated on a PCR thermocycler using the program in (Table 2.6).

RT program		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
42	90:00	1
50	02:00	**10**
42	02:00	
72	15:00	
4	HOLD	HOLD

Table 2.6. RT thermocycler program.

Shortly before the program has finished, thaw and prepare the PCR master mix in a microfuge tube as per (Table 2.7). Once completed, each plate was centrifuged, the PCR master mix added to each plate using the MANTIS and the first PCR step was performed using the program in (Table 2.8). Target-specific primers used for RT and PCR steps were optimised as previously published[122] and are detailed in **Appendix 2**.

PCR mastermix	**1 cell (μL)**	**225 cells + 15% Dead Volume (μL)**
Buffer 2X	6.66	1500
ISPCR (10μM)	0.066	15
SeqAmp (100 U/μL)	0.25	60
cDNA primers (F+R 20μM)	0.037	Variable
gDNA primers (F+R 200μM)	0.11	Variable
Water (DNase/RNase free)	Variable	Variable
TOTAL	**8**	**1800**

Table 2.7. PCR mastermix. F=forward primer, R=reverse primer

PCR program		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
98	3:00	1
98	00:15	24
67	00:20	
72	06:00	
72	05:00	1
4	HOLD	HOLD

Table 2.8. PCR thermocycler program.

After completion of the PCR, the cDNA and amplicon material in each well was diluted with 12μL PCR grade water (cat# AM9935) to a total volume of 25μL. Then 12.5μL from each well was transferred into new 384-well plates as input material for single-cell genotyping, labelled "genotyping stock" (Figure 2.7).

The remaining 12.5μL was kept for full-length transcriptome library preparation. This was performed using a Biomek FxP Automated Liquid Handler platform (Beckman Coulter). Plates at this step were centrifuged, snap frozen and stored at -20C until processed for subsequent steps.

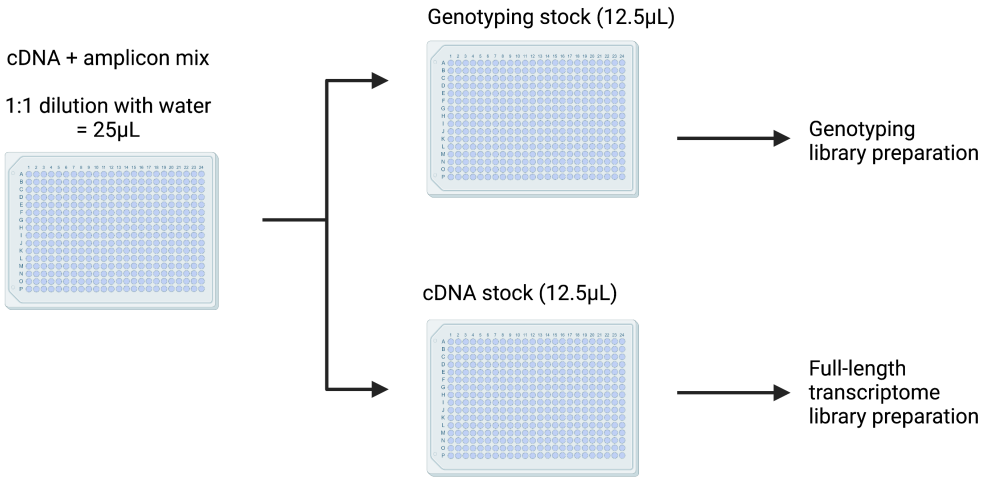


Figure 2.7. Dilution and separate genotyping and transcriptome stock plates. Image generated using BioRender.

2-7.3 High-throughput single-cell genotyping

Targeted single-cell genotyping was performed on the diluted cDNA and amplicon material, "genotyping stock" from the previous steps. Two PCR steps were performed to generate libraries for sequencing on the Illumina platform; PCR1 and PCR2. Both PCR steps were performed using a combination of automated liquid handling platforms (PCR1: Biomek FxP Liquid Handling Platform, PCR2: MANTIS, VIAFLO and Mosquito HTS (TTP LabTech).

The first PCR step, genotyping-PCR1, target-specific primers for gDNA and cDNA amplified regions of interest that were pre-amplified from previous RT-PCR step. These primers have a plate-specific barcode and forward (CS1) and reverse (CS2) adaptor primers (Figure 2.8). The plate-barcode increases the multiplexing potential for sequencing.

Since some of the amplicons for genotyping-PCR1 had different amplification efficiencies, the reactions for each quadrant were split with each reaction contained different primers (Figure 2.9). This is due to various factors including amplification efficiencies of PCR1 primers and expression of gDNA and cDNA amplicons (highly as compared with lowly expressed genes will more easily amplified). Primers used for PCR1 experiments were optimised as previously published[122]. Target-specific primers used for PCR1 genotyping are detailed in **Appendix 3**.



Figure 2.8. Schema of genotyping library preparation. (A) PCR1 step adding primer targeting mutation (single nucleotide variation, SNV, as illustrative example). Primers have 6bp plate barcode (BC1) and adapter (CS1 for forward and CS2 for reverse) for PCR2 primers. (B) PCR2 step adding primer to add well barcode. Primers have CS1/CS2 adapter sequences, Fluidigm well barcode (BC2) illumina sequencing primer adapters (P5 forward, P7 reverse). (C) Final library ready for sequencing. Image generated using BioRender.

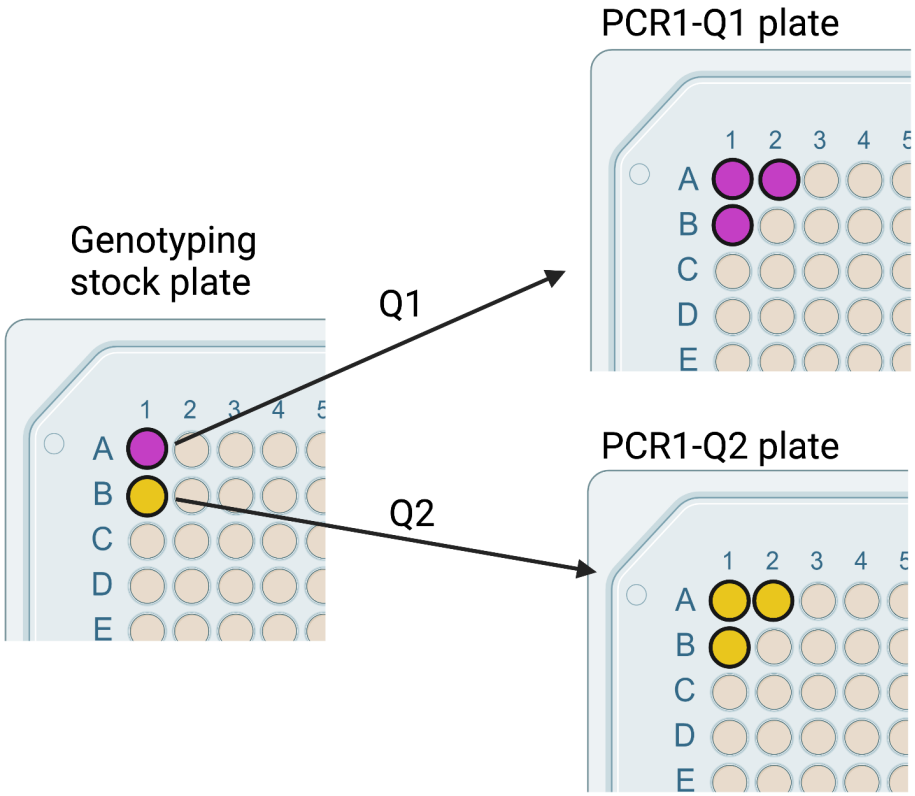


Figure 2.9. Genotyping stock plate transferring to genotyping-PCR1 reaction plate layouts. Image generated using BioRender.

"Genotyping stock" 384-well plates were thawed by spinning down at 1000g for 5 minutes at room temperature to ensure no condensation or droplets on the film when removing to avoid cross-contamination particularly at this stage as the material in each well was still unbarcoded for genotyping. Due to the amplification efficiencies of PCR1 primers, three PCR1 master mixes were prepared (Table 2.9). These master mixes were aliquoted into a 96-well fully skirted plate on a cold block by the Biomek FxP Liquid Handling Platform) which next dispensed 4.75μL of each master mix into the respective quadrant of a new 384-well plate (Figure 2.9) and see TARGET-seq STAR protocol Figure 15[123]) following by adding 1.5μL of genotyping stock material to each well of the new genotyping-PCR1 384-well plate. Once covered with a clear PCR film and spun down, PCR1 program was run on a thermocycler (Table 2.10).

Once the genotyping-PCR1 program was completed, the second PCR step, genotyping-PCR2, was performed (Figure 2.10). In this, a cell-specific Illumina-compatible barcode is added to the

PCR1 library via the CS1 and CS2 adaptors (Figure 2.8B). Illumina-compatible adaptors with cell-specific barcode plates (Access Array Barcode Library for Illumina Sequencers-384, Single Direction; Fluidigm, Cat# 100-4876) were thawed to pre-aliquot 1μL of each 96-well plate into three quadrants of new destination 384-well plates for PCR2.

Genotyping-PCR1 mastermix	**1 cell (μL)**	**110 cells (μL, 1 quadrant)**
KAPA2G Robust HS Ready Mix	3.125	343.75
Primer pair 1 (F+R; 20 μM)	0.188	41.25
Primer pair 2 (F+R; 20 μM)	0.188	41.25
Water (DNase/RNase free)	Variable	Variable
TOTAL	**4.75**	**522.5**

Table 2.9. Genotyping-PCR1 mastermix. Example of one quadrant with 1-2 primers included per reaction depending on primer amplification efficiencies.

Genotyping-PCR1 program		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
95	03:00	1
95	00:15	20
60	00:20	
72	01:00	
72	05:00	1
4	HOLD	HOLD

Table 2.10. Genotyping-PCR1 thermocycler program.

Genotyping-PCR2 Mastermix	**1 reaction (μL)**	**220 reactions (μL; per plate)**
Buffer 10 X (labelled as 3)	0.6	132
MgCl2 (labelled as 4)	1.08	237.6
DMSO (labelled as 5)	0.3	66
Nucl Mix (labelled as 6)	0.12	26.4
Enzyme (labelled as 1)	0.06	13.2
Water (RNAse/DNAse free)	0.84	184.8
TOTAL	**3**	**660**
PCR1 DNA (aliquot with Integra)	1.8	
Barcode (pre-aliquoted with Integra)	1	
TOTAL (cumulative)	**5.8**	

Table 2.11. Genotyping-PCR2 mastermix.

Genotyping-PCR2 master mix was prepared (Table 2.11) and 3μL dispensed per well with the MANTIS in the post-PCR area into a pre-aliquoted barcode new PCR2 plate. After this, 1.8μL of PCR1 product was aliquoted per well into the pre-aliquoted barcode plate with the VIAFLO 96/384 (INTEGRA Biosciences) and the PCR2 program was run on a thermocycler (Table 2.12).

Genotyping-PCR2 program		
Temperature (°C)	**Time**	**Cycles**
95	10:00	1
95	00:15	10
60	00:30	
72	01:00	
72	03:00	1
4	HOLD	HOLD

Table 2.12. Genotyping-PCR2 thermocycler program.

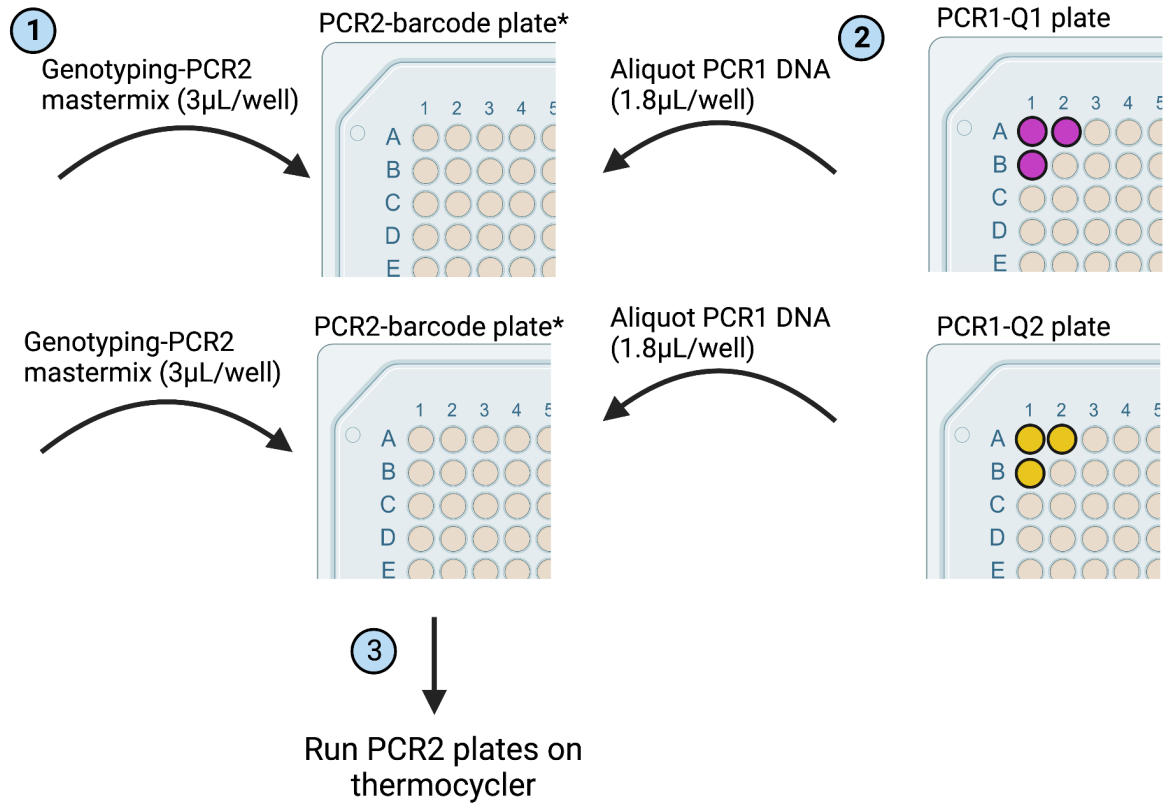


Figure 2.10. Genotyping-PCR2. (1) 3μL of mastermix added to each well of 384-well plate that is pre-aliquoted with 1μL Fluidigm barcodes in each well. (2) PCR1 DNA product (1.8μL) next added to the PCR2-barcode plate containing mastermix which will now total 5.8μL. (3) After spinning down, the PCR2 plates were placed on thermocycler and genotyping-PCR2 program run. Image generated using BioRender.

On completion, 1.5μL from each well of Illumina-compatible barcoded genotyping-PCR2 library quadrant (96 libraries) were pooled using a Mosquito TTP Labtech Liquid Handling Platform. Each pool was cleaned with Ampure XP beads (Beckman Coulter; Cat# A63881) as per manufacturer's instructions at a ratio of 0.8:1 (beads to DNA) ratio adding 80μL of beads to 100μL of pooled genotyping-PCR2 product. The libraries were eluted in 40μL of elution buffer (EB, Qiagen, Cat# 19086).

Library quality was assessed using TapeStation (High Sensitivity D1000 ScreenTape and Reagents, Agilent, Cat# 5067-5583 and Cat# 5067-5584) to check presence of correct library size and distribution (Figure 2.11). Barcodes and adapters added 115bp to the expected amplicon sizes. In parallel, the libraries were quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Cat# Q32854) and then pooled in equimolar amounts to make a final pool of 4nM concentration for sequencing on a NextSeq Illumina instrument. Custom sequencing primers were used to bind to CS1 and CS2 adapters; LCS1 primer for read 1 (15μL of 50μM diluted in 2,485μL of Hybridization Buffer 1 [HT1] buffer), CS2 primer for read 2 (15μL of 50μM diluted in 2,485μL of HT1 buffer) and CS2rc primer for the index read (15μL of 50μM diluted in 2,485μL of HT1 buffer). The final concentration of these primers was 300nM. Custom sequencing primers for genotyping are detailed in Appendix 4. Libraries were sequenced with a configuration: read 1 151 cycles, index read 10 cycles, read 2 151 cycles.

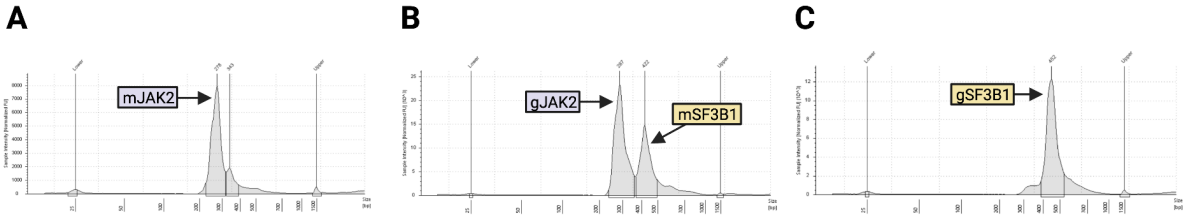


Figure 2.11. Tapestation traces of TARGET-seq genotyping library QC. (A) mRNA/cDNA JAK2 amplicon (287bp peak). (B) Genomic *JAK2* (gJAK2) and mRNA/cDNA *SF3B1* (mSF3B1) amplicons (287 and 422 bp peaks respectively). (C) Genomic *SF3B1* (gSF3B1) amplicon (452 bp peak).

2-7.4 Single-cell full-length transcriptomics

A subset of genotyped HSPCs were selected for full-length transcriptome library preparation which involves sequencing the full-length of the cDNA molecule rather than the 3' or 5' end. At

least 20 cells per HSPC subpopulation (HSC, CD34+CD38- and CD34+CD38+) per genotype in each donor were selected where available. Higher sequencing depth was performed to allow detection of splicing events and therefore study of aberrant splicing in addition to transcriptome information with a depth of 2 million reads selected as detection of splicing events plateaued at 1.5-2 million reads in a prior downsampling analysis (unpublished) of full-length transcriptomic sequencing data from *U2AF1*-mutated MPN HSPCs generated in the Mead lab (Figure 2.12).

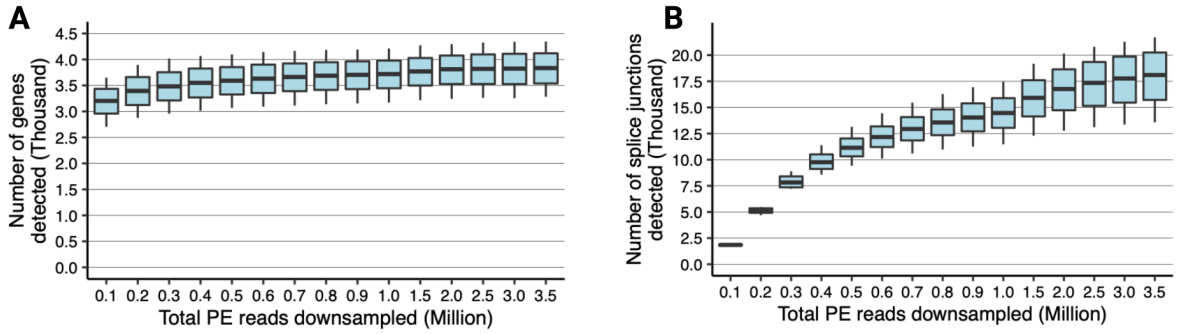


Figure 2.12. Downsampling analysis for splice junction analysis. Boxplots showing number of genes (A) or splice junctions (B) by read depth in a *U2AF1*-mutated MPN dataset. Bioinformatic analysis performed by Wei Wen.

Selected cells from the stock cDNA 384-well plates were transferred and reformatted into new 384-well plates for transcriptome library preparation. This step was performed on the Biomek FxP Automated Liquid Handler platform (Beckman Coulter); a CSV file for each original cDNA plate with the source well location for each cell, the new destination well and the volume to transfer per well (11μL) was required to run the program (Figure 2.13).

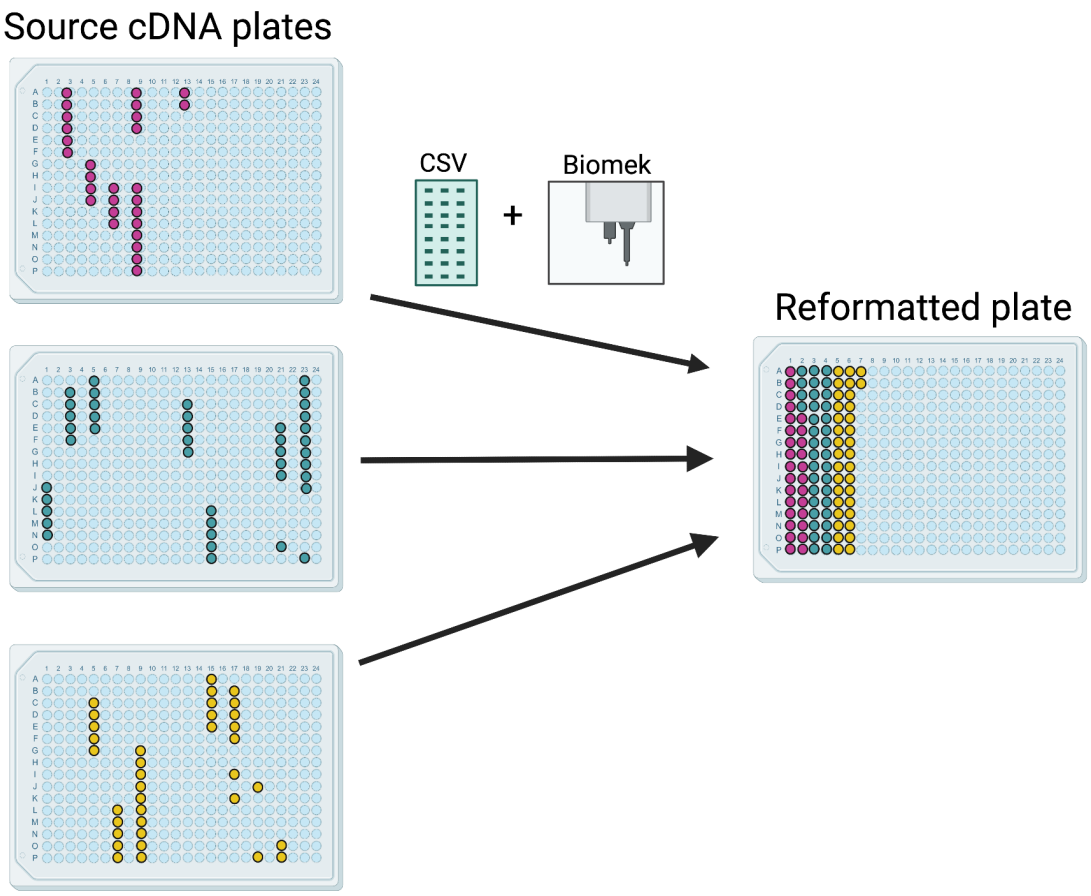


Figure 2.13. Illustrative schema of 384-well plate reformatting from source cDNA plates to destination reformatted plate. Image generated using BioRender.

Full-length transcriptome library was performed on the reformatted 384-well plates in three broad steps (Figure 2.14). First, the cDNA libraries are bead cleaned-up and quality controlled. Next, the libraries are tagmented (fragmented) by a transposase enzyme, Tn5, (Illumina) which has overhang adapters which are tagged on to the ends of the tagmented DNA. Thirdly, each tagmented library is amplified and indexes are added the adaptor tags using i7 and i5 Illumina-compatible primers (Cat# FC-131-2001, Cat# FC-131-2004). This last step is crucial for adding cell barcodes for each well. Until this step, the cDNA of each well was unbarcoded and therefore, library preparation needs to be performed independently in each well to introduce a cell barcode to all of the tagmented molecules. This method was automated with the bead clean-up performed on the Biomek FxP Automated Liquid Handler platform (Beckman Coulter) and the library preparation steps performed using the Mosquito HTS (TTP LabTech).

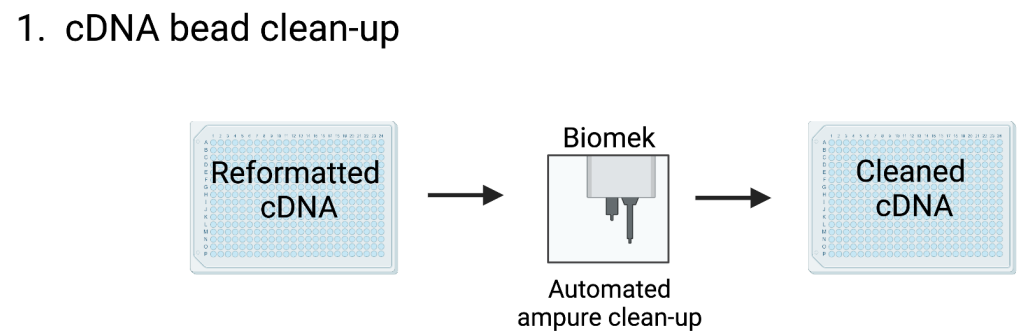


Figure 2.14. Schema of full-length tagmentation method. All steps are automated; Bead clean-up performed with Biomek FxP and tagmentation and indexing steps performing with Mosquito HTS (TTP LabTech). Image generated using BioRender.

Ampure XP bead (Beckman Coulter; Cat# A63882) clean-up steps were performed as per manufacturer's instructions at a ratio of 0.6:1 beads to DNA adding 8μL of beads to 12μL of cDNA. After 5 minutes incubation, this was followed by two 80% ethanol washes and the cleaned cDNA was eluted in 11μL of EB (Qiagen, Cat# 19086). Cleaned cDNA from 48 cells of every 384-well plate were analysed using a Fragment Analyzer (High Sensitivity NGS Fragment Analysis Kit (1–6,000 bp), Advanced Analytical, Cat# DNF-474-1000). The average peak size should be 1.5-2kb.

Nextera XT DNA Library Preparation Kit (96 samples) (cat# FC-131-1096) was used for tagmentation and indexing. The reagents were made up each in Greiner 384 Deep Well plates (cat# 784201) in specific columns; total 4 reagent plates (Table 2.13). Illumina index primers (Nextera XT Index Kit v2 Set A FC-131-2001 and Nextera XT Index Kit v2 Set D FC-131-2004) were aliquoted in sufficient volumes that could be used for at least 4 384-well plates (Figure 2.15). A multihead pipette was used to add 4.5μL of alternating i7 indexes (i.e. N701, N703, N705 etc) into 12 columns of a 384 well Greiner 384 Deep-Well plate starting at column 1 (Plate A i7 index) (Figure 2.15A). Next, 4.5μL of the remaining i7 indexes were aliquoted into 12 columns of a new Greiner 384 Deep-Well plate starting from column 2 (Plate B i7 index) (Figure 2.15B). In a third Greiner 384 Deep-Well plate, 80μL of the i5 indexes were aliquoted into column 1 (Plate C i5 index) (Figure 2.15C).

Reagent	1 cell (μL)	384 cells + 15% dead volume (μL)	Wells/plate
ATM	0.4	11.04	A6 - P6
TD	0.8	22.08	A1 - P1
NT	0.4	11.04	A12 - P12
NPM	1.2	33.12	A1 - P1

Table 2.13. Nextera XT Tagmentation reagents. Each reagent was made up in a separate plate.

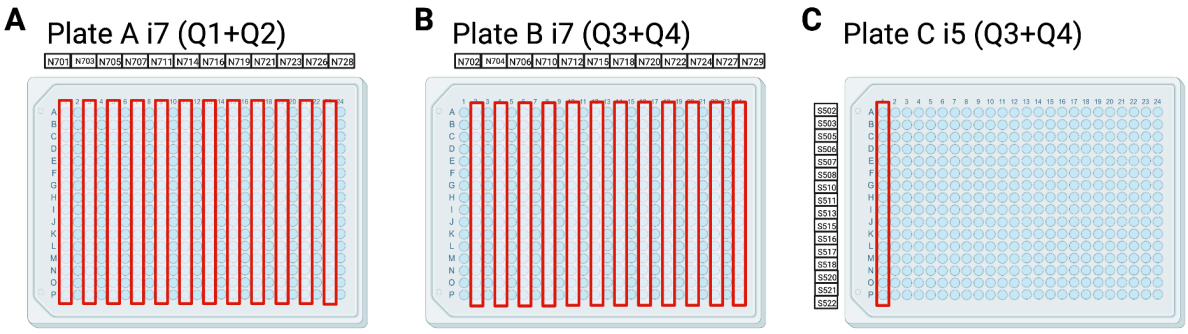


Figure 2.15. i5 and i7 index stock Deep-Well plate layouts.

The following steps were executed using automated programs on the Mosquito. Firstly, 800nL, 400nL, 400nL of TD, cDNA and ATM respectively are aliquoted into each well of a new 384-well plate. Tagmentation was performed on a thermocycler using the program in Table 2.14. The tagmentation reaction was stopped by adding NT buffer for 5 minutes. Next, NPM (1200nL), i5 (400nL), i7 plate A and B indexes were added consecutively to each well.

Tagmentation		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
55	06:00	1
4	HOLD	HOLD

Table 2.14. Tagmentation thermocycler program.

The plate was then spun down and indexing PCR program (Table 2.15) was run on thermocycler. Once the program was finished, 1μL of each well was pooled into one column (A1 to P1) of a Greiner 384 Deep-Well plate. The libraries were manually pooled into one 1.5mL microfuge tube.

Ampure XP bead (Beckman Coulter; Cat# A63882) clean-up was performed as per manufacturer's instructions at a ratio of 0.8:1 beads to DNA adding 160μL of beads to 200μL of DNA. The tagmented libraries were eluted in 150μL of EB (Qiagen, Cat# 19086) and library quality was assessed using TapeStation (High Sensitivity D5000 ScreenTape and Reagents, Agilent, Cat# 5067-5592 and Cat# 5067-5593)) to check presence of correct library size and distribution expecting a broad peak with an average size 300-800bp for tagmented cDNA

Indexing PCR		
Temperature (ºC)	**Time (min:sec)**	**Cycles**
72	03:00	1
95	00:30	1
95	00:10	10 cycles
55	00:30	
72	00:30	
72	03:00	1
4	HOLD	HOLD

Table 2.15. Indexing PCR thermocycler program.

Libraries were quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Cat# Q32854) and adjusted to make a final pool of 4nM concentration for sequencing on a NovoSeq Illumina instrument with standard Illumina sequencing primers. Libraries were sequenced with the following configuration: 151 cycles for read 1, 10 cycles for index read and 151 cycles for read 2.

2-8 Statistical analysis of clinico-molecular datasets

Individual mutations and mutation groups (epigenetic, splicing factor, high molecular risk) were correlated with clinical characteristics and outcomes using Pearson's chi-squared test, Student's t-test and Mann-Whitney U test. Univariate and multivariate logistic regression was applied to examine the associations between mutations and clinical outcomes (treatment responses, survival outcomes and adverse events). Adjustment was performed for multivariate analyses if any correlate was statistically significant ($p<0.05$) on univariate analysis or clinically relevant in the analysis. Adverse events were included in this analysis if grade 3 or 4 (classified by the National Cancer Institute Common Terminology Criteria for Adverse Events v4.03). Median follow-up time calculated using the reverse Kaplan-Meier method[142] with the event of interest reversed. Survival outcomes were performed where relevant and included overall survival, transformation-free survival (TFS) and thrombotic/haemorrhagic-free survival using Kaplan-Meier method and log-rank test. TFS analysis was applicable to the MAJIC-ET cohort only with myelofibrosis, leukaemia or death considered events for TFS calculation. Multivariate analysis of survival outcomes was performed using the Cox proportional hazards regression model; including in the model covariates which had a p-value <0.05 on univariate cox regression analysis. Statistical analyses were performed using SPSS version 25. StataCorp. 2017. and R statistical software package version 3.4.0[143] with RStudio version 1.1.463.

Individual plasma cytokine levels were correlated with specific somatic gene mutations and clinical demographic data followed by an unsupervised approach which applied hierarchical agglomerative clustering (*hclust* function in stats R package, with "ward.D2" method) to identify related sets of cytokines. A cluster score was derived based on the median overall cytokine concentration within a cluster for each patient. Scores were then correlated clinico-genomic data was analysed. Statistical analyses were performed using SPSS version 26 and R statistical software package version 3.6.0 with RStudio version 1.1.463. to each cytokine. Cluster scores were correlated with mutation and clinical information.

2-9 Bioinformatic analysis of 10X/CITE-seq data

2-9.1 Droplet-based single-cell RNA-sequencing and CITE-seq pre-processing

RNA-sequencing libraries were pre-processed to generate gene expression counts per cell (gene expression matrix format) using 10X genomics analysis pipeline, Cell Ranger. The data were processed with Cell Ranger version 7.0.0 which had additional features of including intronic mapped reads (account for 20-40% of the reads) and antisense reads which were not included in earlier Cell Ranger versions. Both intronic and antisense reads correlate with sense reads suggesting these are derived from poly-A tracts and are not generated from genomic DNA. Including these recovers more usable reads and yields higher UMIs and therefore genes per cells.

First, the Illumina sequencing output files, BCL files were demultiplexed to Fastq files using "cellranger mkfastq". Using "cellranger count", Fastq files were then aligned to the human reference genome, GRCh38 (refdata-cellranger-GRCh38-3.0.0), then filtered with cell barcodes and UMIs counted to generate gene expression counts per cell (gene expression matrix format).

For CITE-seq library processing, the Illumina sequencing BCL files were also demultiplexed to Fastq files using "cellranger mkfastq". Fastq files were then processed using CITE-seq-Count (version 1.4.4). The Fastq file for read 1 contained the cell barcode and UMI and the Fastq file for read 2 contained the antibody barcode. "CITE-seq-Count" command also required a CSV file with a list of the antibody barcodes required and expected cell count. The expected cell count included was higher than the number of cells sorted to incorporate empty droplets containing barcodes and UMIs but no cells to downstream set the background threshold of UMI reads for each antibody above which correspond to a positive antibody signal.

2-9.2 Donor demultiplexing, quality control and normalisation

Cell Ranger output files (gene expression matrix) were next demultiplexed by donor and quality control filtering was performed. Donor demultiplexing was performed using a combination of methodologies. Firstly, I incorporated hash tag oligonucleotides (HTO) which were included at cell sample staining step.[144] These are oligonucleotide-tagged antibodies against a surface protein expressed on all cells and therefore could be used to label different donors. The Illumina

sequencing output BCL files for HTO libraries were processed in the same manner as for CITE-seq using both “cellranger mkfastq” and “CITE-seq-Count”.

Doublets are an important confounder to exclude in single cell transcriptomics. These are typically classed into homotypic, where the cells are transcriptionally alike (either from the same donor or cell type), or heterotypic doublets where the cells are transcriptionally distinct (either distinct cell types or different donors)[145]. Doublet detection methodologies can be experimental and computational. I incorporated two doublet exclusion methodologies into my experimental design noting that both experimental and computational methods are more sensitive for heterotypic than homotypic doublet detection. The first method, Cell Hashing[144], involved labelling the cells with oligo-tagged antibodies. These antibodies are reactive to proteins expressed universally in all cells and therefore the hashing label can uniquely stain the cells of each donor facilitating pooling of different donors / cell types for sequencing with downstream demultiplexing by detection of the unique oligonucleotide tag / barcode. The second methodology, Souporcell[146], leverages the donor genetic SNP signatures from the RNA-sequencing reads allowing for cross-donor doublet detection and genetic donor demultiplexing. Souporcell was run on input files (*possorted_genome_bam.bam* and *barcodes.tsv*) with the following 6 steps: 1) remapping was performed with *minimap2*, 2) candidate variants were called using *freebayes*, 3) cell allele counting with *VarTrix*, 4) cells were clustered by genotype (souporcell.py), 5) doublets were called using *troublet* and 6) cluster genotypes were called and ambient RNA estimated “*consensus.py”*. HTO and Souporcell were compared and the method providing a superior cell assignment and doublet exclusion. After cell assignment, QC filtering was performed using the following thresholds: UMI reads ≥1000, number of genes per cell >500, mitochondrial read <15% per cell, genes expressed in at least 10 cells and singlets as determined by HTO and/or Souporcell.

Using the Seurat 4.3.0 package, the libraries were combined and next, to remove the influence of sequencing depth of each cell, RNA-sequencing reads were normalised using *NormalizeData* function applying *LogNormalize* method. The gene counts per cell were divided by the total counts per cell, multiplied by a scale factor of 10,000 and finally log-transformed. The most highly variable genes (n=2000) were selected with *FindVariableGenes* function for downstream analysis to improve the biological signals yielded[147]. Linear transformation called ‘scaling’ was performed on the data with *ScaleData* function prior to applying dimensionality reduction methods. CITE-seq/ADT libraries were also normalised to mitigate the effect of sequencing depth. Centred log

ratio transformation (CLR)[140] was the normalisation method applied with *NormalizeData* function setting "margin" to 2 to perform normalisation across cells rather than ADT marker.

2-9.3 Dimensionality reduction, data integration and clustering of 10X genomics data

PCA was performed on merged libraries prior to integration using the Seurat function *RunPCA* and UMAP dimensionality reduction was implemented using Seurat function *RunUMAP* with the top 30 PCA dimensions. Two integration methods were compared to determine the optimal integration strategy; harmony[148] and reciprocal principal component analysis (RPCA)[147]. Harmony was performed on RNA-sequencing data using the Seurat function *RunHarmony* on the merged libraries with the following parameters chosen: group.by.vars = "sort", kmeans_init_nstart=20, kmeans_init_iter_max=100 and theta=2. UMAP dimensionality reduction was repeated using top 30 harmony dimensions. Unsupervised clustering was performed using a graph based KNN clustering approach with harmony dimensions. First, the graph was constructed by running the Seurat *FindNeighbors* function using harmony top 30 dimensions as input. To then cluster the cells, the Seurat function *FindClusters* was run testing a range of resolution parameters (0.8 – 1.2) to determine the optimal resolution to define cell clusters.

RPCA was performed in parallel which identifies shared features ("anchors") between the samples and projects these features onto a shared PCA space. Normalisation steps and selecting the most highly variable genes was performed as previously described but carried out for each library individually rather than on combined data. Recurrently variable genes between libraries were identified performing the Seurat function *SelectIntegrationFeatures.* Each library was then individually scaled and PCA dimensions identified using *RunPCA.* "Anchors" were identified using the Seurat function *FindIntegrationAnchors* taking the list of libraries as input, using the recurrently variable genes between the libraries as the "anchor.features" and setting the reduction to run to "rpca". The Seurat function *IntegrateData* was run using these "anchors" to integrate the libraries. The integrated data was next scaled and PCA dimensions determined. Ensuring the assay was set to "integrated", the same steps as for Harmony integration from performing UMAP to cell clustering was performed.

Local Inverse Simpson's Index (LISI[148]) was the computation method used to objectively compare harmony and RPCA integration methods generating a score for each in addition to visual

inspection. LISI was performed using *compute_lisi* from the LISI package[149] which requires UMAP coordinates of each integration methods and cell type and library batch as covariates to compute cLISI and iLISI scores respectively.

Cell population annotation was performed on the cell clusters for each individual library and after integration of all libraries. The final cell cluster annotation was determined after comparing the marker genes expressed by each cluster, the lineage marker gene signatures and the expression of canonical lineage ADT markers by each cluster.

Multimodal integration using RNA and ADT information was also performed and compared with RNA integration alone. Weighted nearest neighbour analysis (WNN) from the Seurat package was performed. Harmony integrated data as described previously was used as input for this. ADT was independently integrated using the same Harmony steps as described previously. The Seurat function *FindMultiModalNeighbors* was performed using the following options: reduction.list = list("harmony", "harmony.adt"), dims.list = list(1:30, 1:40).

2-9.4 Lineage trajectory analysis

Trajectory analysis was performed to reconstruct HSPC lineage trajectories in the 10X dataset. Force-directed graph layout was performed using ForceAtlas2 (FA2) software. This was applied using SingCellaR package's *runFA2_ForceDirectedGraph* function. A second trajectory analysis method was performed, Monocle3 pseudotime analysis, with SingCellaR package. The Monocle object was generated with *new_cell_data_set* function. Harmony UMAP and FDG coordinates were used to plot pseudotime results with *plot_cells* function.

2-9.5 VarTrix

Genotype calling was performed on 10X genomics RNA-sequencing data using the software, VarTrix (https://github.com/10xgenomics/vartrix). The input files required for this software were *possorted_genome_bam.bam* and *barcodes.tsv*. The scoring method (--scoring-method) used was *consensus* which outputs to a matrix file a score 0 for no reads at the variant position, 1 WT reads, 2 for mutant reads and 3 for both WT and mutant reads.

2-9.6 Differential abundance analysis

Differential abundance (DA) analysis aims to identify differences in proportions between two groups. DA analysis was performed using MiloR[150]. This method is agnostic of cell clustering and instead assigns cells to overlapping neighbourhoods on a k-nearest neighbour (KNN) graph which has in cell populations that differentiate along a continuum such as HSPCs and thus are often difficult to assign to accurate distinct cell clusters. A Milo object was created using the function *Milo.* A KNN graph was constructed using the function *buildGraph*. Neighbourhoods were determined using the function *makeNhoods.* Differential abundance testing was performed using the function *testNhoods* which incorporates experimental co-variates to test (e.g. donor type) and technical covariates can be included. Batch corrected harmony dimensions were used as input for building the KNN graph mitigating the need to include this co-variate in the DA test stage.

2-9.7 Differential gene expression analysis

Differential gene expression analysis (DGEA) was performed using the Seurat function *FindMarkers*. The non-parametric Wilcoxon test was used on log2 normalised values to compare the mean of the gene expressed between two groups of cells. Significance was set for genes as follows: adjusted p-value <0.1, log2(fold-change) >0.5 and the gene expressed in ≥ 25% of the cells in either of the groups. P-value adjustment was performed using Bonferroni correction.

Gene-set enrichment analysis (GSEA) was performed using SingCellaR package[149] (https://github.com/supattlab/SingCellaR). Gene sets were downloaded from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/ http://software.broadinstitute.org/gsea). Firstly, a pre-ranked gene list was created by performing DGEA for two cell populations and then the log2 fold change of each gene was multiplied by -log10 of the adjusted p-value. This was performed using the function *identifyGSEAPrerankedGenes.* GSEA was carried out using the function *Run_fGSEA_analysis* with the following input data: pre-ranked gene and the query gene-set gmt file and the normalised enrichment scores (NES) plotted.

2-9.8 Regulatory network analysis

A regulatory network, or regulon, consists of transcription factors and their associated downstream target genes. Single Cell rEgulatory Network Inference and Clustering (SCENIC)

enables this at the single cells level. First, targets of TFs were identified based on co-expression building a gene regulatory network (GRN) from the gene expression matrix as the input file using *pyscenic grn* command. Possible targets that bind directly to TFs were determined using DNA-motif analysis using *pyscenic ctx* command. The regulons in each cell were scored using AUCell to facilitate analysis of specific regulon enrichment.

2-9.9 Single-cell genotyping analysis of plate-based

Single-cell genotyping libraries were demultiplexed to fastq files in two steps; first, running bcl2fastq (version 2.20, Illumina) to demultiplex by 384-well position followed by demultiplexing by plate using fastq-grep to demultiplex by plate barcode and fastq-pair to pair reads, both programs from fastq-tools (version 0.8). Up to six plate (PCR1) barcodes were used for sequencing pooling increasing the multiplexing capacity to 384 cells times equating to up to 2304 cells sequenced per run.

Fastq files are pre-processed to call genotype for each cell using a custom perl script.[123] This required a tsv file of the variants Reads were aligned to the human genome (hg19) using STAR with default settings (version 2.4.2a). mRNA and gDNA amplicons were separated and variants called using mpileup program in samtools (version 1.1) and summarised per cell into a text file.

2-9.10 Reconstruction of clonal hierarchies in TARGET-seq genotyping data

Prior to assigning genotypes per cell, the limit of detection was set. Reads in the blank wells were used to set this for each amplicon; the highest value in the blank plus 50 reads were added to obtain the limit of detection for each variant location.

Both *JAK2*V617F and *SF3B1* mutations were SNVs. For both mRNA and gDNA reads, the number of reads from the alternative allele were divided by the reads from the reference allele. Depending on the proportions detected, cells were assigned WT, heterozygous or homozygous. Variant allele frequency (VAF) was determined as the fraction of alternative to reference reads and expressed as a percentage. Genotypes for assigned for mRNA and gDNA based on the following VAF proportions; <2% WT, 4-98% heterozygous and 99-100% homozygous Genotypes for mRNA and gDNA were merged to determine a final genotype.

2-9.11 TARGET-seq full-length transcriptomic pre-processing and analysis

Nucleotide bases from the 3'-end of the reads with Phred score <20 were trimmed using TrimGalore (version 0.6.5)[151]. Trimmed reads were mapped to the human reference genome (GRCh38 build) using STAR aligner (version 2.6.1d) in 2-pass mode [152]. In the 1st pass, the reads were mapped to the genome to detect the splice junctions in each individual cell [*--outSAMtype None*]. In the 2nd pass, the splice junctions detected in the 1st pass across all cells were used as references to re-detect the splice junctions in each individual cell. This approach will rescue low-expressing splice junctions missed in the 1st pass. In the 2nd pass, the binary alignment mapped (BAM) file was also generated for each individual cell *[--outSAMtype BAM SortedByCoordinate --outSAMattributes NH HI AS nM XS --quantMode TranscriptomeSAM –limitSjdbInsertNsj*]. The splice junction counts were collated into a matrix whereby the rows represent splice junction coordinates while the columns represent cell IDs.

Gene expression was quantified using featureCounts under the subread module (version 2.0.3) with the default parameters[153]. The gene expression counts were collated into a matrix whereby the rows represent gene IDs while the columns represent cell IDs. For each cell, the total gene counts were computed, and then the gene expression counts for each cell were normalised using their respective total gene counts. The normalised expression values were subsequently scaled using the mean total gene counts across all cells (1,278,357).

Next, sequencing depth, alignment rate, and percentage mitochondrial reads were subsequently computed. To this end, sequencing depth was retrieved from the log files of TrimGalore. Total number of mapped reads and mitochondrial reads were computed using Samtools (version 1.9) *stats* and *view*, respectively [154]. Using these values, the alignment rate was computed by dividing total number of mapped reads over sequencing depth and then multiplied by 100. Percentage mitochondrial reads was computed by dividing total number of mitochondrial reads over sequencing depth and then multiplied by 100.

Sequencing depth, alignment rate, percentage mitochondrial reads, number of genes detected, and number of splice junctions detected were used as QC metrices to classify each cell as passed or failed QC. A gene is considered to be expressed when the normalised expression value is ≥1 while a splice junction is considered to be expressed when the raw count is ≥1. Cells with

sequencing depth >25,000, percentage alignment >25, percentage mitochondrial reads <25, >100 expressed genes, and >100 expressed splice junctions were included for downstream analysis. In total, 1,167 of 2,304 (70%) of cells passed QC.

The functions performed for this analysis were derived from SingCellaR package[149] (https://github.com/supattlab/SingCellaR) unless otherwise stated. Normalisation was performed using "normalize_UMIs" function: gene count/library size) x median library size of the dataset. Highly variable genes were selected using "*get_variable_genes_by_fitting_GLM_model*" which applies a generalised linear model (GLM) of log2 of the mean expression level and coefficient of variation of each gene. PCA was run using "*runPCA*" function followed by "*runUMAP*". Batches were integrated using Harmony with "*runHarmony*" function with donor effect regressed by including as covariate. Lineage differentiation trajectory analysis was performed using force directed graph (FDG) with the function "*runFA2_ForceDirectedGraph"* and diffusion map with the function "*runDiffusionMap*". Pseudotime analysis was performed along trajectories using monocle (version 3_0.2.3.0). TARGET-seq data was projected onto 10X HSPC clusters using Seurat functions *FindTransferAnchors, TransferData* and *MapQuery.*

2-9.12 Splicing event detection and quantification

The BAM files for each donor were merged using Samtools *merge*. A catalogue of all transcripts detected in each donor was generated using StringTie (version 2.1.4) with the default parameters[155]. The resulting gene transfer file (GTF) from each donor was merged using StringTie to generate a catalogue of all transcripts detected in this study. Custom R scripts were used to format the GTF so that novel transcripts will be detected in the downstream splicing event detection tools.

The splicing events skipped-exon (SE), mutually exclusive exons (MXE), retained intron (RI), alternative 5' and 3' splice sites (A5SS and A3SS) were detected using rMATS version 4.1.0 [156]. The per-base coverage of the introns of each RI event was computed using BEDTools (version 2.71.1) *coverage* [*-split -d*][157]. The sum of the per-base coverage of the introns of each RI event was then computed and collated into a matrix whereby the rows represent the intron coordinates while the columns represent the cell IDs.

Additional splicing event detection and quantification were performed using MARVEL[158]. The splice events detected by rMATS, intron count matrix, normalised gene expression matrix, splice junction count matrix, and GENCODE v31 GTF were used to create the MARVEL object using the *CreateMarvelObject* function. Additional splicing events, namely alternative first and last exons (AFE and ALE) were detected using the *DetectEvents* function. Next, each splicing event corresponding to SE, MXE, RI, A5SS, A3SS, AFE, and ALE was quantified in percent spliced-in (PSI) unit using the *ComputePSI* function. A splice event is considered to be expressed when it is supported by ≥10 splice junction reads. The PSI values of splicing events supported by <10 splice junction reads were re-coded as missing values (NA's).

Additional pre-processing was performed prior to downstream analysis with MARVEL. Firstly, the normalised gene expression values were offset by +1 and then log2-transformed using the *TransformExpValues* function. Any value less than 1 was re-coded as 0. Secondly, any cell IDs coded as "blank" or "batch_control" in the sample metadata were excluded. The former cells represent blank wells while the latter cells represent the same cells used across all 386-well plates during TARGET-seq library preparation. Furthermore, only cells with genotype information were retained.

2-9.13 Differential splicing and gene expression analysis

Differential splicing analysis was performed between two pre-defined cell groups using the *CompareValues* function. A splicing event was included for differential analysis when it is expressed in ≥20 cells in both cell groups. Anderson-darling and D Test Statistics (DTS) were used to assess the differences in PSI distribution between the two cell groups. *P* values were adjusted using false discovery rate (FDR). A splicing event was considered differentially spliced when $P < 0.1$ and $\Delta PSI > 10$. Genes of differentially spliced events were included for differential gene expression analysis using the *CompareValues* function. Wilcox test was used to assess the differences in median gene expression values between the two cell groups. *P* values were adjusted using FDR. A gene was considered differentially expressed when $P < 0.1$ and log2 fold change > 0.5.

Chapter 3: The molecular landscape of high-risk MPN

3-1 Introduction

Therapeutic application of JAK2 inhibitor therapy in MPNs swiftly followed the identification of *JAK2*V617F mutations in ~50% of MPNs[10,11,159]. RUX, the first-in-class and clinically approved JAK1/JAK2 inhibitor, was approved in 2012 for MF following the success of COMFORT-1 and 2 trials and is now standard for care for management of symptomatic patients with splenomegaly and constitutional symptoms[133]. Moreover, RUX is FDA approved for the second line management of PV[160].

Although RUX has been associated with clinical responses, molecular responses have been limited with no disease modification in most MF patients. Treatment with JAK inhibitors in early MPN phases (e.g. in PV/ET) and in specific molecular subgroups of these may demonstrate disease modification and alter the natural disease course. To explore the therapeutic role for RUX in ET patients, the phase 2 MAJIC trial randomised patients with hydroxycarbamide resistance/intolerance (HC-RES/INT ET) 1:1 to receive RUX or best available therapy (BAT). The primary outcome was achievement of ELN[161] complete haematological response (CHR) within one year[162]. The primary analysis showed equivalent rates of CHR in both treatment groups[136] but mutational profiles were not extensively reported in this analysis. In ET, mutations in non-MPN driver genes (NDM) have been reported in (29 - 72%)[163-165] with the most common affecting *TET2*, followed by *ASXL1*, *DNMT3A* and *SF3B1* and critically, certain mutations are associated with an inferior prognosis[164]. However, NDM analysis in ET has yet to be widely implemented in clinical practice and important areas of uncertainty have remained relating to the impact of NDM in particular subgroups such as HC-RES/INT ET, the effect of RUX treatment on disease course in molecularly-defined subgroups and whether serial NDM assessment has value.

MF is a more advanced MPN that is biologically and clinically heterogeneous with certain subgroups of patients more challenging to treat[1]. NDM are detected more often than in ET/PV; in >90% of patients, with particular mutations important prognostically; specifically mutations in epigenetic (ASXL1, EZH2)[166], splicing factor (SRSF2, U2AF1)[167], *TP53*[112], *RAS* pathway[168] and IDH1/IDH2 genes[169] are associated with disease progression and a shortened survival. Although

RUX has improved outcomes with reduced splenomegaly and increased survival, a significant proportion of patients fail RUX treatment due to intolerance or poor disease response. Discontinuation rates of RUX have been used as a surrogate for RUX failure as a clear definition is lacking with rates varying across studies from 50-89% at 3-years[74]. Poorer RUX response has been associated with ≥3 mutations[170] but not with mutation type[171,172]. Shorter time to RUX failure has been reported in those with ASXL1/EZH2 mutations[171] and reduced time to RUX discontinuation with ≥3 mutations[170].

RUX-associated anaemia and thrombocytopenia are common in the first three months of treatment but thereafter tend to improve. However, RUX-associated thrombocytopenia is associated with adverse outcomes[173,174]; those who discontinue RUX in association with thrombocytopenia have a median survival is less than 1 year[174].

RUX is not licensed for use in patients with severe thrombocytopenia[175.176] but pacritinib (PAC), a JAK2/IRAK1/FLT3/CSF1R inhibitor, has shown clinical benefit in thrombocytopenic myelofibrosis patients initially in the PERSIST-1 and -2 trials[177,178] and most recently in the PAC203 study, a randomized dose finding study of PAC in patients with primary or secondary myelofibrosis who were RUX-resistant/intolerant (RES/INT), including patients with severe thrombocytopenia. This study established the optimal efficacious and safe dose[179] for PAC which is now FDA approved for treatment of thrombocytopenic MF. The molecular profiles of RUX-RES/INT thrombocytopenic MF patients have not been well delineated and better understanding of their mutation profiles will aid the application of precision medicine for this challenging group of patients with a major unmet need for new management strategies.

3-1.1 Hypothesis and aims

In MPN, I hypothesise that there is an enrichment of specific molecular subgroups associated with resistance to treatment in both earlier and more advanced MPN stages. Moreover, these molecularly-defined groups may have an associated characteristic clinical and laboratory (including cytokine) profiles. In earlier MPN (e.g. ET), I hypothesise that treatment with JAK inhibitors may modify the natural history of disease transformation in specific molecular groups of patients. I sought to evaluate the molecular profiles of treatment-resistant MPN to examine the

molecular characteristics associated with treatment resistance, clonal evolution after prolonged treatment and correlate these with clinical outcomes.

Specifically:

1. To evaluate the molecular profiles associated with treatment resistant MPN (ET and MF)
2. To correlate molecular profiles with clinical and laboratory characteristics
3. To correlate molecular profiles with outcomes, specifically disease progression and treatment response

To achieve these aims, I interrogated two clinical trial MPN patient cohorts; HC-RES/INT ET on the MAJIC-ET trial and RUX-RES/INT on the PAC203 trial, which importantly have well annotated clinical information to enable correlative analysis between mutation profiles and clinical information such as patient characteristics and outcomes such as treatment responses. After describing a distinct mutation profile within one of these cohorts, I interrogated independent clinical trial datasets inclusive of a published dataset to understand whether this profile was enriched in MPNs.

3-2 Results

3-2.1 The mutation profile of high-risk HC-RES/INT ET patients is distinct from low-risk ET

Mutational analysis was performed on 110 patients randomized to BAT (47.3%, n=52) or RUX (52.7%, n=58) on the MAJIC-ET phase 2 trial with median follow-up time of 55 months (95% CI 49.9 – 60.4). *JAK2*V617F, *CALR* and *MPL* mutations were present in 49.1%, 30% & 4.5% of patients, respectively, with no driver mutation detected in 16.4% of patients, denoted “triple-negative” (TN). Baseline NDM were present in 30% (n=33) of patients, most frequently *TET2* (n=12) and a particular enrichment for *TP53* (n=7) and *SF3B1* (n=7) mutations (Figure 3.1A; Table 3.1, Appendix 5).

Baseline Clinical Characteristics		**Patients with NDM n (%)**	**Patients without NDM n (%)**	***P***
All patients (n=110)		33 (30)	77 (70)	
Median age in years (range)		71 (44 – 91)	64 (35 – 85)	0.0001
Gender	**Female** **Male**	17 (51.5) 16 (48.5)	49 (63.6) 28 (36.4)	0.234
HC-Resistant HC-Intolerant		19 (57.6) 14 (42.4)	34 (44.2) 43 (55.8)	0.197
Treatment	**BAT** **RUX**	17 (51.5) 16 (48.5)	35 (45.5) 42 (54.5)	0.560
Disease Duration at TE (y)		6.6 (0.8 – 31)	7.7 (0.4 – 25.9)	0.699
No. of Prior Therapy Lines	**<3** **≥3**	27 (81.8) 6 (18.2)	52 (67.5) 25 (32.5)	0.127
Previous Thrombosis Previous Haemorrhage		7 (21.2) 2 (6.1)	28 (36.4) 4 (5.2)	0.118 0.855
Baseline blood counts (median, range)	**WCC (x 10^9/l)** **Hb (g/l)** **Plt (x10^9/l)**	5.8 (1.7 – 15.2) 115 (90 – 147) 517 (166 – 1406)	6.1 (2.6 – 29.8) 125 (87 – 160) 530 (89 – 1139)	0.917 0.01 0.927
Driver mutation status	**J*AK2*V617F** ***CALR*** ***MPL*** **TN**	19 (57.6) 7 (21.2) 3 (9.1) 4 (12.1)	36 (46.8) 26 (33.8) 2 (2.6) 13 (16.9)	0. 237
*JAK2*V617F allele burden ≥50%		5/19 (26.3)	6/36 (16.7)	0.395
CALR allele burden ≥50%		0/7 (0)	2/26 (7.7)	1.0

Table 3.1. Baseline clinical characteristics in patients with and without non-MPN driver mutations (NDM). Differences were analysed using the chi-squared test for categorical variables or fisher's exact test if a value less than 5 in any cells of the contingency table and non-parametric Mann-Whitney U test for continuous variables (i.e. age, disease duration, blood counts). MPN=myeloproliferative; patients=patients; n=number; BAT=best available therapy; Hb=haemoglobin; HC=hydroxycarbamide; NPM=non-MPN (myeloproliferative neoplasm) driver mutation; P32=radioactive phosphorus; Plt: platelet count; RUX=ruxolitinib; y=years; TN=triple negative; WCC=white cell count.

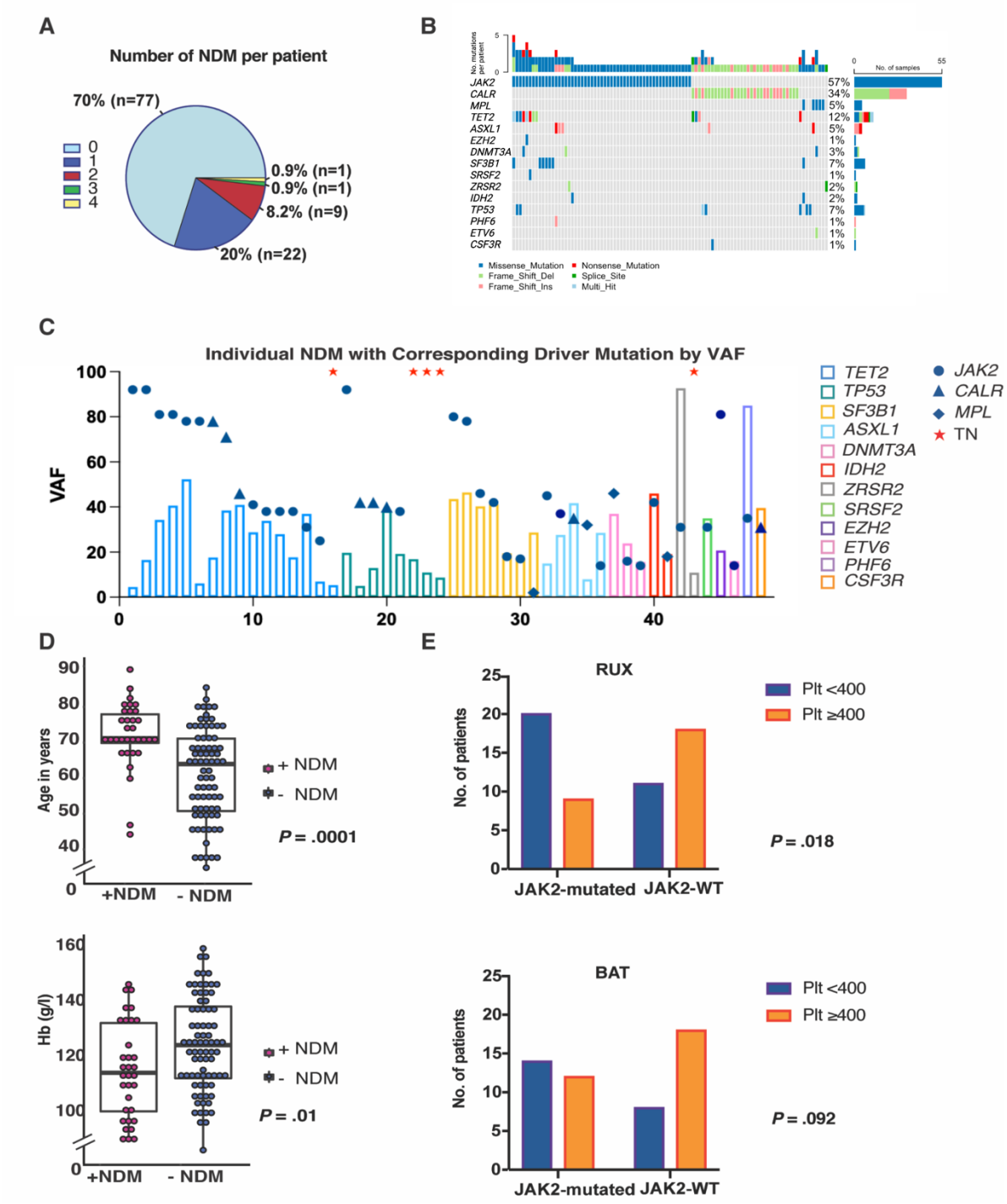


Figure 3.1. Baseline mutational analysis and correlation with clinical characteristics and outcomes. (A) Pie chart showing number of NDM per patient. (B) Waterfall plot of mutation distribution in the MAJIC-ET cohort. (C) Column and dot plot showing variant allele frequencies (VAF) of each NDM (column) with corresponding driver mutation (blue dot) listed in descending order of driver mutation VAF. Red star indicating TN patient; driver mutation VAF was higher in 66.67%, 87.5% and 20% of *JAK2*, *CALR* and *MPL*-mutated patients suggesting driver mutation acquisition first in these. (D) Dot and box plots of median age in patients with NDM compared to patients without NDM; 71 vs. 64 years, p=0.0001 (upper plot) and

haemoglobin (Hb) levels (mean Hb 115g/l) lower in patients with NDM compared to patients without NDM (median Hb 125g/l), p=0.01 (lower plot). Dots represent each individual patient and each horizontal line and box represent the median for age/mean for Hb and interquartile ranges respectively using Mann-Whitney U test to compare median ages (non-normal distribution) and Student's t-test to compare Hb means (normal distribution). (E) Post hoc analysis of 1-year platelet count responses; significantly more patients on RUX who were *JAK2*-mutated achieved plt ≤400 than non *JAK2*-mutated patients (upper bar chart). This difference was not seen within the BAT arm (lower bar chart). BAT=best available therapy; *JAK2*=*JAK2*V617F; NDM=non-MPN driver mutation; Plt ≤400=platelet count of ≤400 x 10^9/l; Plt >400=platelet count of > 400 x 10^9/l; RUX=ruxolitinib; TN=Triple negative.

The latter two mutations are present at higher frequencies than previously reported in ET cohorts[43,163-165]. One NDM was present in 20% (n=22), two in 8.2% (n=9) and one patient each with 3 and 4 NDM (Figure 3.1B). Driver mutation variant allele frequency (VAF) was higher in 66.67%, 87.5% and 20% of *JAK2*V617F, *CALR* and *MPL*-mutated patients respectively (Figure 3.1C) consistent with driver mutations being the disease-initiating, first acquired mutations. However, it is important to note that the order of mutations can only be definitively proven using single-cell methodologies given confounding factors such as loss of heterozygosity at the JAK2 locus[180]. Neither driver mutation status or VAF correlated with presence of NDM (p=0.237 and p=0.582 respectively). However, *TP53* mutations showed a trend to higher frequency in TN (17.6%) than in *JAK2*V617F*/CALR/MPL*-mutated patients (4.3%), p=0.073. NDM were present equally across CALR mutation subtypes.

Patients with NDM tended to be older at trial entry; 71 years vs. 64 years without NDM, p=0.0001 (Table 3.1; Figure 3.1D). They also had lower haemoglobin (Hb) at TE than patients without NDM; Hb 115g/l versus Hb 125g/l, p=0.01 (Table 3.1; Figure 3.1D). Lower Hb specifically correlated with presence of a splicing factor (SF) mutation, p=0.014.

Longer history of disease duration at TE was not associated with the presence of NDM overall. However, adjusting for age, *TP53*-mutated patients had longer disease duration at TE (median 11.02 years, range 6.3 - 31.2) versus *TP53*-wild type patients (median 6.8 years, range 0.4 – 38.9), p=0.027. *SF3B1* mutations were more prevalent in HC-resistant patients (11.3%, n=6) than HC-intolerant patients (1.8%, n=1), p=0.054. Furthermore, more *SF3B1*-mutated patients had baseline palpable splenomegaly (42.9%, n=3/7) vs. wild type (WT) patients (6.8%, n=7/103), p=0.017 and more often had prior haemorrhage (28.6%, n=2/7) than WT patients (3.9%, n=4/103), p=0.046.

3-2.2 Correlation of mutation profiles with clinical responses and outcomes

i. *JAK2*V617F-mutated patients experienced significantly more platelet responses in ruxolitinib-treated patients

In the primary analysis, no correlation was detected between driver mutation status and complete haematological response (CHR) defined by platelet count <400 and normal spleen size.[136] As platelet count reduction is a key therapeutic goal in ET, we carried out a post-hoc analysis of platelet count response defined as a platelet count <400 x 10^9/l at 1-year. *JAK2*V617F-mutated patients had significantly more platelet count responses compared to non- *JAK2*V617F-mutated patients (Figure 3.1E).

Outcome (n_E)	**OR (UV)**	**OR 95% CI**	***P***
1-year CR (ELN) (n_E=50)	0.72	0.3 – 1.6	0.43
1-year post hoc platelet response (n_E =55)	0.5	0.2 – 1.2	0.12
1-year Overall symptom score response (≥50% reduction) (n_E =12)	0.71	0.2 – 2.6	0.61
Transformations (MF n_E =13, AML n_E =1)	3.8(NDM) (adj)* 14.4 (HMR) (adj)*	0.7– 21.5 1.7-122.8	0.129 0.015
Thrombotic event (n_E =21)	0.75	0.2 – 2.3	0.62
Haemorrhagic event (n_E =10)	2.1 (NDM)(adj) 18.9 (SF)(adj)	0.5 – 10 2.2 - 161	0.34 0.007
Death (n_E =13)	2.14	0.6– 2.3	0.21
Stopping RUX treatment (n_E =40)	0.38 (adj)	0.7 – 1.9	0.25
Grade 3/4 haematological toxicity (anaemia/thrombocytopenia) (n_E =17)	1.5 (adj)	0.4– 5.9	0.56

Table 3.2. Logistic regression predicting the influence of non-MPN driver mutations (NDM) on clinical outcomes. All models were adjusted for *JAK2*V617F mutation status and treatment type since patients were stratified by these at trial entry. Further adjusted analysis was performed for outcomes with significant (p<0.05) odd ratios (OR) on initial analysis to include age, TE haemoglobin level and platelet counts denoted indicated with "(adj)" next to OR. The presence of a HMR mutation significantly increased

the odds of a transformation event, p=0.015. Haemorrhagic events outcomes were associated specifically with the presence of SF mutations. adj=adjusted; AML=acute myeloid leukaemia; CHR=complete haematological response; CI=confidence interval; ELN; European LeukaemiaNet; HMR=high molecular risk mutation (SF and/or *TP53* mutations); MF=myelofibrosis; n_E =number of events; non-MPN (myeloproliferative neoplasm) driver mutation; OR=odds ratio; RUX=ruxolitinib; SF= splicing factor mutation (*SF3B1, ZRSR2, SRSF2)*; TE=trial entry; UV=univariate. *Driver mutation allele burden ≥50% and gender also included in final model for transformation.

This was specific for RUX-treated patients; 64.5% in *JAK2V617F*-mutated vs. 35.5% in non *JAK2V617F*-mutated, p=0.018. The presence of NDM at baseline did not influence haematological responses or 1-year symptom responses (Table 3.2). RUX discontinuation was correlated with non *JAK2V617F*-mutated status (OR 6.1 [95% CI 1.43 – 26.6], p=0.015) but not presence of NDM (OR 0.38 [95% CI 0.7 – 1.2], p=0.25, Table 3.2).

ii. Transformation events in MAJIC-ET were specifically associated with *TP53* and splicing factor mutations with no amelioration of this risk with ruxolitinib treatment

Overall survival (OS) at 4-years was 91.5% (95% CI 80.2– 96.4%) in BAT and 83% (95% CI 70.4% - 90.5%) in RUX-treated patients (log-rank test, p=0.22). OS was not affected by driver mutation status or NDM. Transformation events occurred in 12.7% (14/110) of patients during the period analysed. 93% (13/14) of these were progression to myelofibrosis with one transformation to AML. *JAK2*V617F-mutated patients had a poorer 4-year TFS (69.3%, 95% CI 54.1 – 80.3%) than non-*JAK2*V617F-mutated patients (85.3% 95% CI 71.2 – 92.8%), p=0.045 (Figure 3.2A). Driver mutation allele burden (≥50%) was associated with a poorer 4-year TFS 43% (95% CI 20.9 – 88.3% vs. TFS 81.1% (95% CI 72.3 – 91%) than for patients with VAF <50%, p=0.005. The presence of NDM at baseline was predictive of an inferior 4-year TFS of 65.4% (95% CI 46.2 – 79.1%) vs. 82.8% (95% CI 70.9 – 90.2%), in patients without NDM (p=0.017; Figure 3.2B). *TET2* mutations were not associated with a statistically significantly inferior TFS (p=0.084) but mutated-*TP53* was associated with an inferior 4-year TFS; *TP53*-mutated patients 42.9% (95% CI 9.8 – 73.4%) vs. *TP53*-wild type (WT) patients 79.8% (95% CI 69.7 – 86.8%), p=0.011 (Figure 3.2C). SF mutations also conferred a poorer 4-year TFS of 40% (95% CI 12.3 – 67%) vs. 81.5% for WT patients (95% CI 71.4– 88.3%; p=0.00039; Figure 3.2D).

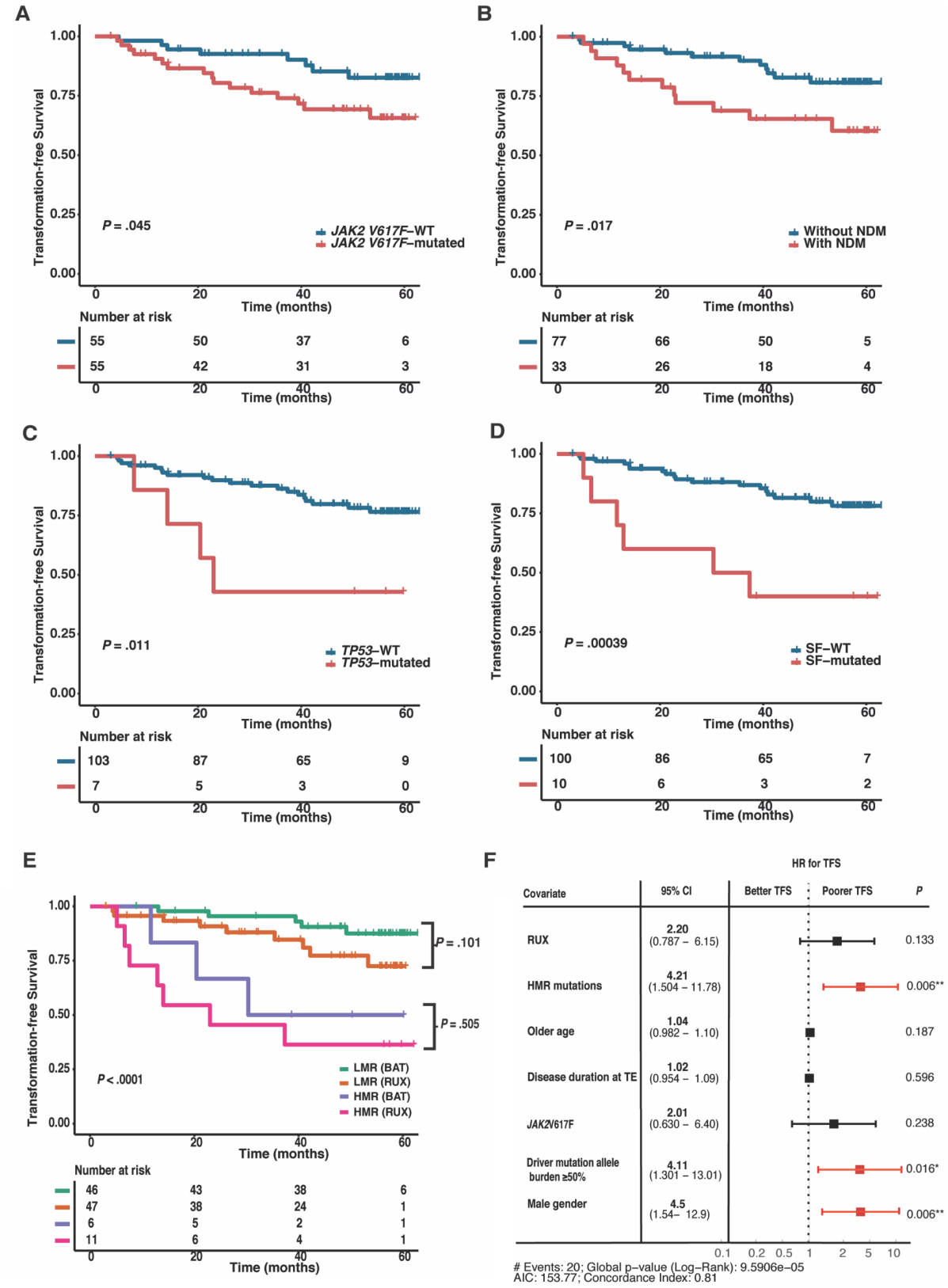


Figure 3.2. Kaplan-Meier curves of transformation-free survival (TFS) stratified by mutational statuses with survival estimates, reported at 4-years. (A) *JAK2*V617F-mutated patients had a poorer

4-year TFS (69.3% [95% CI 54.1 – 80.3%]) when compared to non-*JAK2*V617F-mutated patients (85.3% [95% CI 71.2– 92.8%]), p=0.045. (B) NDM at baseline was predictive of an inferior 4-year TFS; patients with NDM (65.4% [95% CI 46.2 – 79.1%]) vs. patients without NDM (82.8% [95% CI 70.9 – 90.2%]), p=0.017. (C) *TP53* mutations were associated with an inferior 4-year TFS; *TP53*-mutated (42.9% [95% CI 9.8 – 73.4%]) vs. *TP53*-wild type (WT) patients (79.8% [95% CI 69.7 – 86.8%]), p=0.011. (D) SF mutations conferred a poorer 4-year TFS; SF-mutated (40% [95% CI 12.3 – 67%]) vs. SF-WT (81.5% [95% CI 71.4 – 88.3%]), p=0.00039. (E) Stratifying patients with high risk molecular (HMR) mutations in this study by treatment arm demonstrates no amelioration of negative impact with RUX treatment; patients with HMR on RUX had TFS at 4-years of 36.4% (95% CI 26.2 – 46.6%) and on BAT 50% (29.1 – 67.7%) (p=0.505 between these arms) as compared to those without these mutations (i.e. low molecular risk, LMR) with TFS at 4-years of 84.7% (95% CI 71.6 – 92%) on RUX and of 90.6% (95% CI 78.5 – 96%) on BAT (p=0.101 between these arms). (F) Forest plot showing multivariable cox model of transformation-free survival (TFS). Covariates significant on univariate analysis were included. Age and disease duration at TE were categorized as continuous variables. The log-rank test was used to compare survival estimates between groups. CI=confidence interval; HMR=high molecular risk; LMR=low molecular risk; *JAK2*=*JAK2*V617F; NDM=non-MPN

This was predominantly attributable to *SF3B1* mutations (p=0.004). I combined SF and *TP53* mutations to define a HMR (high molecular risk (HMR) group and compared the effect of RUX treatment on TFS in these patients. Importantly, RUX treatment did not ameliorate their adverse effect on TFS (Figure 3.2E).

Multivariable analysis of TFS (Figure 3.2F) including treatment type, HMR mutations, age, disease duration at TE, *JAK2*V617F status, driver mutation allele burden ≥50% and gender showed that the presence of HMR mutations retained their negative impact on TFS independently; HR 4.21, 95% CI 1.5 – 11.8, p=0.006. RUX treatment did not mitigate the poor prognostic impact of these mutations (HR 2.2, p=0.133). Driver mutation allele burden ≥50% retained negative impact independently in this model (HR 4.11, p=0.016) and male gender was also associated with an inferior TFS (HR 4.5, p=0.006). This has been reported in other ET cohorts[43,108,181] but the cause remains elusive.

Thrombotic and haemorrhagic events (CTCAE [Common Terminology Criteria for Adverse Events] version 4.0 grades 1-5) occurred in 19.1% (21/110) and 9.1% (10/110, respectively; 47% (10/21) and 30% (3/10) respectively were > grade 2. Thrombotic events were not influenced by either driver mutation status (OR 2.4 [95% CI 0.87 – 6.4], p=0.09) or NDM (OR 0.75 [95% CI 0.2 – 2.3], p=0.62). Similarly, driver mutation and NDM status did not impact thrombosis-free survival.

Driver mutation status did not influence haemorrhagic events (OR 0.65, 95% CI 0.2 – 2.4, p=0.505) or haemorrhagic-free survival. The presence of NDM was not associated with increased risk of a haemorrhagic event (OR 2.1 [95% CI 0.5 – 10], p=0.34) but patients with SF mutations did more often experience a haemorrhagic event; OR 18.9, 95% CI 2.2 – 161, p=0.007 (Table 3.2). At 4-years, haemorrhagic-free survival for SF-mutated patients was 19% (95% CI 13.4 – 25.2) vs. 86.2% (95% CI 78.2 – 91.3%) for SF-WT patients, p<0.0001. In this subgroup, trial entry platelet counts ranged 221 – 704 x 10^9/l with platelet counts closest to the time of haemorrhagic event normal in 3 of these patients (*SF3B1*-mutated; 318 & 397 x 10^9/l, *SRSF2*-mutated 373 x 109/l) and reduced (*ZRSR2*-mutated; 54 x 10^9/l) and elevated (*SF3B1*-mutated, 504 x 10^9/l) in one patient each, respectively. Grade 3/4 haematological toxicities were experienced by 15.5% (17/110) of patients; 94.1% (16/17) in RUX-treated patients and these were not associated with the presence of driver mutations or NDM.

3-2.3 Longitudinal mutation analysis in MAJIC-ET patients

i. Molecular responses were rare in this cohort and confined to RUX-treated patients

Patients with known driver mutations (85%, n=79/93) underwent driver mutation molecular analysis at 12 months. Driver mutation molecular responses (MR) at 12 months were rare (n=3) and exclusively seen in the RUX arm; a complete molecular response (CMR) in 2 patients *(JAK2*V617F-mutated and *CALR*-mutated) and a partial molecular response (PMR) in one *CALR*-mutated patient. In patients with driver mutations, 54% (50/93) had a sample available for >12 month time point allele burden analysis. Median sample time was 48 months after TE (24 – 60 months). There was no significant change in allele burden in the overall cohort at 12 months (median change 15.3%, 0-400%) or >12 month time point (median change 21.6%, 0-389%) (Figure 3.3A-B).

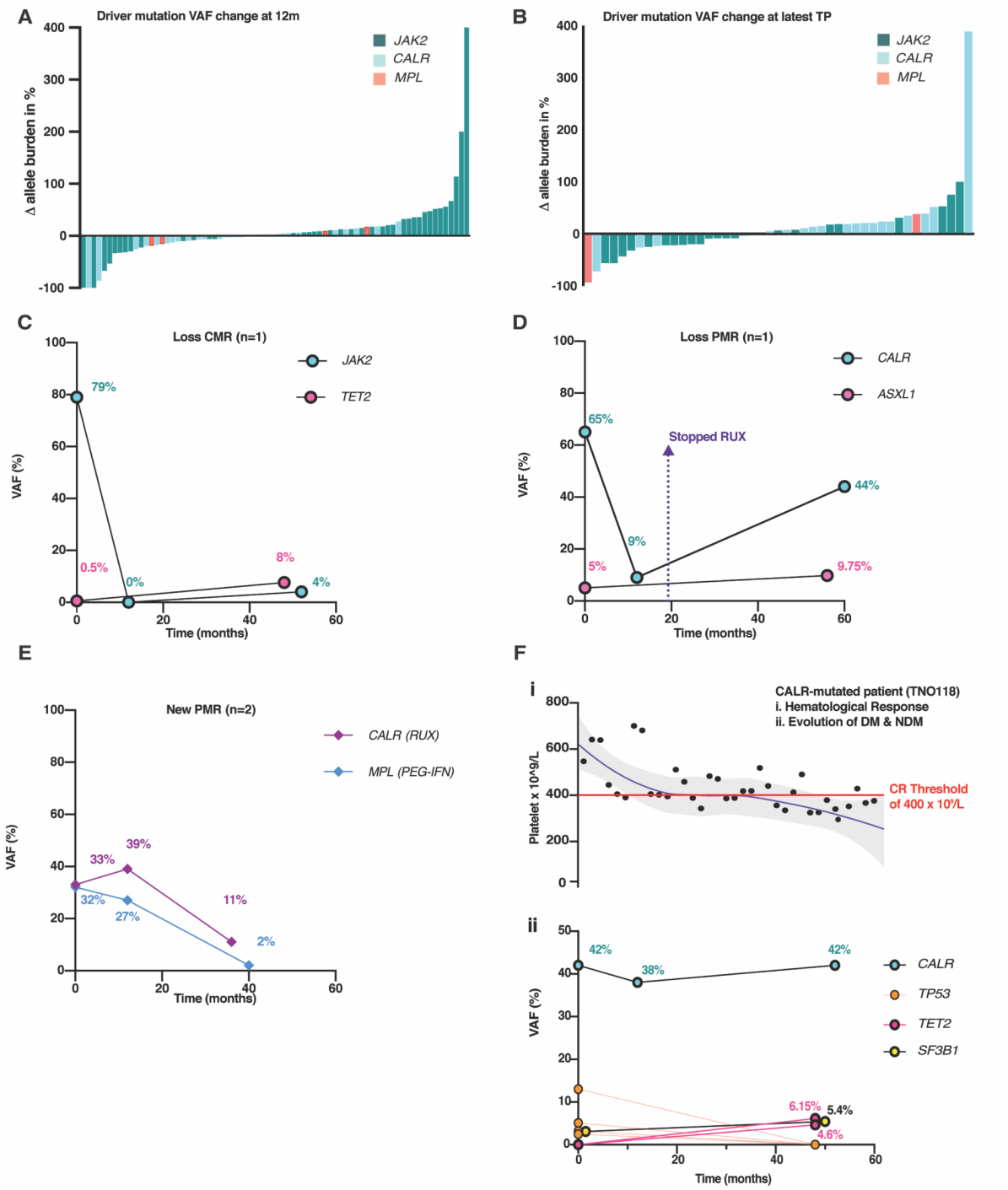


Figure 3.3. MAJIC-ET longitudinal mutational analysis. (A) Waterfall plot of driver mutation change in VAF for each patient at 12months; median change 15.3% (0-400%) and (B) driver mutation change in VAF from 12m to latest time point; median 21.6% (0-389%). (C) BAT-treated patient achieving a *JAK2 V617F* CMR at 12 months; VAF 22 to 0%. Subsequently, they had a loss of CMR at 44 months with *JAK2 V617F* VAF 4% coinciding emergence of low level *TET2* mutation. (D) RUX-treated patient achieving a *CALR* PMR at 12 months; VAF 65 to 9%. Subsequently, they had a loss of PMR with a *CALR* VAF 44% at 60 months and antecedent to this, an *ASXL1* mutation emerged at 56 months. Notably, this patient switched from RUX

at 20 months due to toxicity. (E) New PMR in two patients; a *CALR*-mutated patient on RUX and the second a *MPL*-mutated patient on pegylated-interferon (VAF 27 to 2%). (F) Dot plots illustrating a *CALR*-mutated pt on RUX with platelet counts (i) and change in mutation VAFs (ii) during the study period; co-mutated *TP53* (n=4) at TE and at most recent follow up, disappearance of these mutations with emergence of a new *SF3B1* (VAF 5.44%) and *TET2* (n=2; VAF 4.66 & 6.15%) mutations. BAT=best available therapy; CMR=complete molecular response; VAF=variant allele frequency; RUX=ruxolitinib; PMR=partial molecular response; TE=trial entry; n=number.

For the 3 patients with MR at 12m, 2 of these subsequently lost MR; a *JAK2V617F*-mutated patient lost CMR at 44 months (*JAK2V617F* VAF of 4%) (Figure 3.3C) and a *CALR*-mutated patient lost PMR at 60 months (*CALR* VAF of 44%) (Figure 3.3D). New PMRs occurred in 2 patients; a *MPL*-mutated patient on pegylated-interferon and a *CALR*-mutated patient on RUX (Figure 3.3E). Although not fulfilling PMR as per ELN MR criteria, 4 *JAK2V617F*-mutated patients with a baseline allele burden of <20% had a ≥50% reduction at the most recent time point sample; RUX-treated (n=2) and interferon-treated (n=2).

ii. Clonal evolution was associated with loss of molecular response

Later time point NDM analysis was carried out in 52% of the cohort (57/110) with a median sample time of 40 months after TE (6 – 60 months), including 25 of 33 patients with baseline NDM. 91.7% of baseline NDM persisted at later time point samples with only 8.3% of mutations undetectable at follow-up; 2 *TP53* mutations (TE VAF 5.09% and 13.01%), both in the same CALR-mutated patient treated with RUX (Figure 3.3F). This patient achieved CHR by 12 months and the platelet count continued to remain close to CHR throughout the study period (Figure 3F). The third NDM not detectable at follow-up was in a TN patient treated with BAT (anagrelide); a low level *TET2* mutation was no longer detectable at 40 months with a concomitant baseline *TP53* mutation that persisted (VAF reduced from 11% to 3%). One patient (TN) had an over 4-fold increase in *TP53* mutation allele burden from 9 to 38%, corresponding to time of leukemic transformation.

New NDM were frequently acquired (16 mutations in 10 patients (10/57, 17.5%) (Appendix 5); *TET2* [5], *ASXL1* [4], *TP53* [3], *SF3B1* [2], *U2AF1* [1], *SETBP1* [1]. However, 87.5% (14/16) of these mutations were detected below the pre-specified threshold for variant calling (5%) on review of baseline sequencing annotation files indicating that the true rate of new NDM is only 3.5%. These mutations were equally acquired in *JAK2*V617F- and *CALR*-mutated patients with no mutations acquired in *MPL*-mutated or TN patients. New mutations occurred at a similar

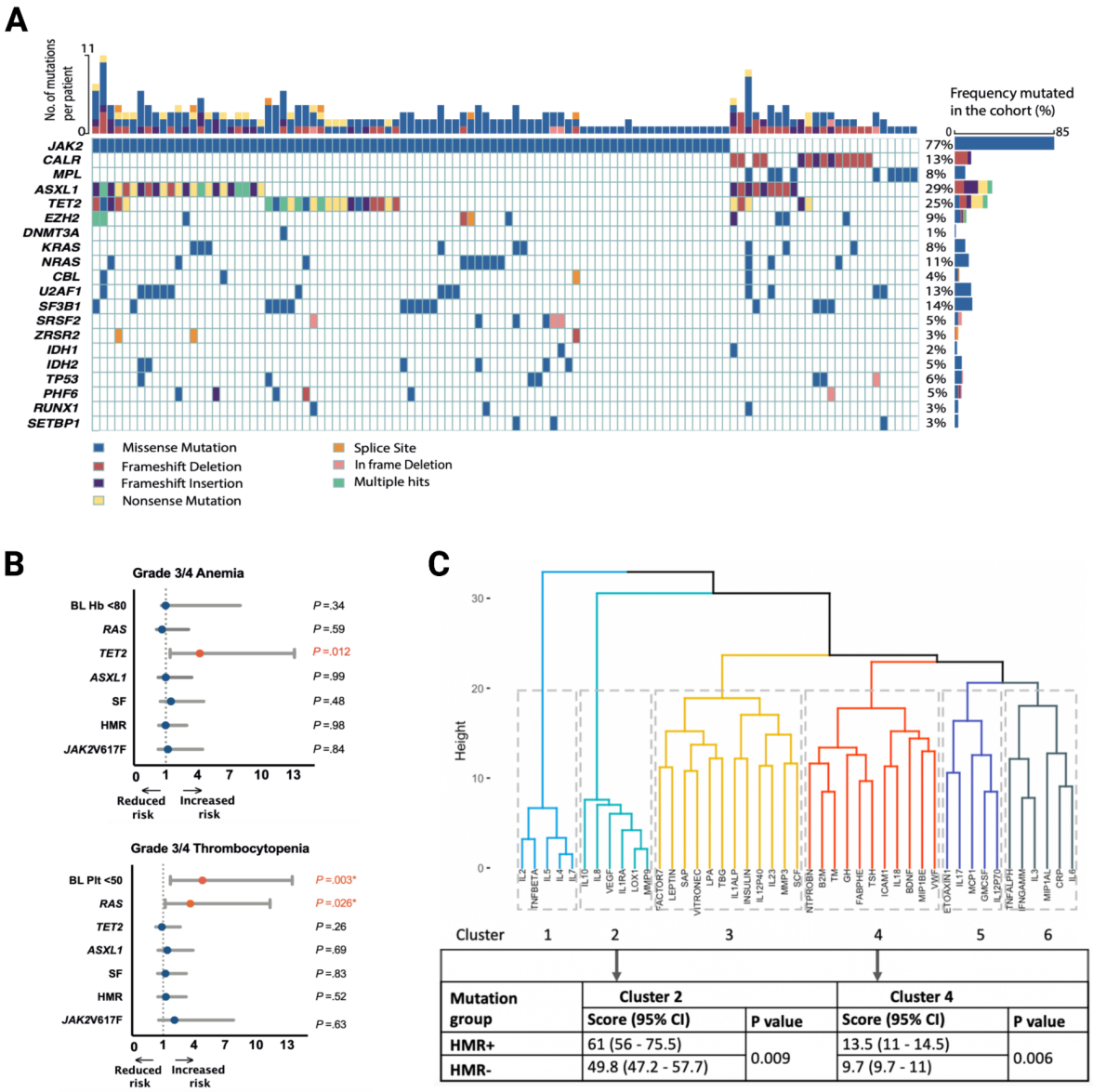
frequency in RUX (6/10) and BAT (4/10) arms, p=0.74. Only one patient with disease progression and later time point available sample had a new NDM; a low level *TET2* (VAF of 6.79%) at 12 months. There was no association between acquisition of a new NDM and presence of NDM at baseline, clinical characteristics or outcomes. However, only a median follow-up time of 10.7 months (95% CI 9.05 – 12.4%) was available after later time point NDM analysis which is an insufficient duration for transformation and survival analysis.

Notably, both the above patients who lost driver mutation MR had evidence of clonal evolution with emergence of a new *TET2* mutation in the *JAK2V617F*-mutated patient with loss of CMR (Figure 3.3C) and emergence of an *ASXL1* mutation in the *CALR*-mutated patient with loss of PMR (Figure 3.3D). Two BAT-treated patients, both *JAK2V617F*- and *TET2*- co-mutated at baseline, showed emergence of more than one new NDM; 3 and 4 NDM respectively.

The key findings from this serial molecular analysis were that MPN driver mutation molecular responses were rare during the time period analysed (up to 5 years) and patients who lost MPN driver mutation molecular responses had associated emergent NDM. Further, serial profiling of NDM identified a low rate of true new NDM and there was no correlation with disease progression.

3-2.4 Distinct mutational profile in RUX-RES/INT MF patients treated with JAK2 inhibition

To examine mutation profiles associated with RUX-RES/INT MF, I analysed the PAC203 patient cohort who were enriched for thrombocytopenia. Patients were randomised 1:1:1 (PAC 100mg once daily [QD], 100mg twice daily [BID] or 200mg BID) and stratified by baseline platelet count. 110 patients (of 164 recruited) on the PAC203 study underwent mutation profiling. Clinical characteristics of this subgroup of patients was representative of the overall PAC203 cohort[179]. Median follow-up time was 213 (95% CI 189 - 236) days. The median prior exposure to RUX was 1.59 years (range 0 – 11 years) and 72.7% had prior exposure to non-RUX therapies (range 1 – 5 lines). The median age was 68 (37-87) years. Over half of the cohort were male; 58.2% (64/110) and 41.8% female (46/110). Primary myelofibrosis (PMF) was the most prevalent disease category (56.4%, 62/110), followed by post-polycythaemia vera MF (PPV-MF) in 29.1% (32/110) and post-essential thrombocythaemia MF (PET-MF) in 14.5% (16/110). Thrombocytopenia was common; the median baseline platelet count was 63 $x10^9$/L, with <50 $x10^9$/L in 38.2% (42/110). Baseline Hb was <100g/L in 64.5% (71/110) of the cohort.



Mutation group	Cluster 2		Cluster 4	
	Score (95% CI)	P value	Score (95% CI)	P value
HMR+	61 (56 - 75.5)	0.009	13.5 (11 - 14.5)	0.006
HMR-	49.8 (47.2 - 57.7)		9.7 (9.7 - 11)	

Figure 3.4. Mutational landscape and clinical correlations in the PAC203 cohort. (A) Waterfall plot of mutation distribution in the PAC203 cohort. (B). Forest plot illustrating the results of logistic regression analyses of mutation statuses and baseline haemoglobin count associated with the likelihood of grade 3/4 anaemia; *TET2*-mutated patients were more likely to experience grade 3/4 anaemia independent of baseline haemoglobin level; odds ratio (OR) 4.5, 95% CI 1.4 – 13.9, *P*=.009 (upper panel). Forest plot illustrating the results of logistic regression analyses of mutation statuses and baseline platelet count associated with the likelihood of grade 3/4 thrombocytopenia; *K/NRAS*-mutated patients were more likely to experience a grade 3/4 anaemia AE after adjustment for baseline platelet level; OR 3.65, 95% CI 1.2 – 11.3, *P*=.026**.** (lower panel). Univariate logistic regression was performed for each variable. Significant P values (<0.05) highlighted in red and OR denoted by (*) were adjusted. JAK2V617F=JAK2 V617F-mutated; HMR=high molecular risk mutation [*IDH1/2, SRSF2, ASXL1, EZH2, U2AF1*Q157]; SF=splicing factor mutation [*SF3B1*, *U2AF1*, *SRSF2*, *ZRSR2*]; *ASXL1= ASXL1*-mutated; *TET2=TET2*-mutated; *RAS=K/NRAS*-mutated; BL Plt <50=baseline platelet level <50 x10^9/L. **(C)** Cluster dendrogram of cytokine

levels in ruxolitinib refractory / intolerant study cohort with cluster scores 2 and 4 highlighted in the table for high molecular risk positive (HMR+) and negative (HMR-) patients.

*JAK2*V617F mutations were present in 77.3% (85/110), *CALR*-mutated (12.7%, 14/110; type 1 n=11, type 2 n=3), *MPL*-mutated in 8.2% and "triple-negative" (1.8%) cases, Figure 1A. *JAK2*V617F variant allele frequency (VAF) was ≥50% in 68.2% (58/85) with <20% VAF present in just 5.9% (n=5) of patients. NDM were present in 76.4% (n=84) with ≥3 NDMs in 20.9% (23/110) of patients. The most prevalent NDMs were *ASXL1* and *TET2* (29.1%, n=32, and 25.5%, n=28, respectively), Figure 3.4A.

Splicing factor (SF) mutations were mutually exclusive and detected in 34.5% (n=38/110) of patients: *SF3B1* [13.6%, n=14], *U2AF1* [12.7%, n=14], *SRSF2* [5.5%, n=6], *ZRSR2* [2.7%, n=3]. Patients with SF mutations were most often categorised as PMF (76.3%) rather than PET-MF (15.8%) or PPV-MF (7.9%), *P*=.0001, as previously reported[182,183]. More SF-mutated patients were male (72.2%) vs. in SF-wild type (51.4%), *P*=.047. SF-mutated patients had lower baseline haemoglobin levels (Hb <8g/dL in 39.5% as compared with 16.9% in SF-WT, *P*=.009) and were more likely to be red cell transfusion dependent at trial entry (RCC-D) as compared with SF-WT patients (43.2% vs 22.5% respectively, *P*=.012). *SF3B1*-mutated patients had higher trial entry platelet counts (platelet count >100x10^9/L) in 66.7% vs. 28.7% in *SF3B1*-WT patients, *P*=.004).

High molecular risk mutations (HMR; *IDH1/2, SRSF2, ASXL1, EZH2, U2AF1*Q157)[167] were present in 43.6% (48/110) of patients and ≥2 HMR mutations were present in 14.5%, a prevalence similar other high-risk enriched MF cohorts[182,184]. No clinical parameters were associated with a HMR mutation. Strikingly, RAS-pathway activating mutations, *KRAS/NRAS/CBL* (*RAS/CBL*-MT), were found at a higher frequency than previously described in MF cohorts[168,184]; 20.9% (n=23; *RAS* n=20, *CBL* n=3). These mutations were sub-clonal in the majority with a median VAF 10% (range 1.9 – 95%). *RAS* mutations occurred in known mutation hotspots; the most prevalent in G12 codon (n=12/20)[170]. RAS/CBL-MT patients were more likely to have ≥3 NDM (52.2% vs. 12.6% for RAS-WT patients, *P*=.0001) and a co-mutated HMR mutation (65.2% vs 37.9% for RAS-WT patients, *P*=.019), Table 3.3. *KRAS/NRAS/CBL* and *TP53* (n=7) mutations were mutually exclusive in this cohort.

	***RAS*-mutated, N=20 n (%)**	***RAS*-WT, N=90 n (%)**	***P* value**
Age in years (median, IQR)	69 (65.25, 75.0)	67.5 (62.75, 73.0)	0.447
Male gender, n (%)	11 (55)	53 (58.9)	0.75
MF diagnosis, n (%) Primary MF PPV MF PET MF	 11 (55) 7 (35) 2 (10)	 51 (56.7) 25 (27.8) 14 (15.6)	 0.723
Prior ruxolitinib, n (%) Failure Intolerance	 15 (75) 15 (75)	 65 (73) 62 (68.9)	 0.857 0.59
Prior ruxolitinib duration in months (median, IQR)	30.2 (13.8, 56.6)	16.5 (6.3, 41.1)	0.06
Platelet count x 10^9/L (median, IQR)	60 (31, 118.5)	63 (40, 135)	0.549
Platelet count <50, x 10^9/L, n (%)	9 (45)	33 (37.1)	0.511
Haemoglobin <10g/dL, n (%)	16 (80)	55 (61.1)	0.123
RBC transfusion status, n (%) Dependent	7 (35)	25 (28.4)	0.796
Platelet transfusion dependent n (%)	3 (15)	6 (6.7)	0.225
Peripheral blasts (median, IQR)	2 (1.5, 2.5)	2 (0.2, 3)	0.405
White blood cells, $\times 10^9$/L (median, IQR)	6.36 (3.17, 18.4)	7.02 (4.16, 15.5)	0.572
Spleen Volume (cm^3) by MRI/CT (median, IQR)	2576 (1327, 3782)	2274 (1599, 3051)	0.348
Driver mutation status *JAK2 V617F* *CALR* *MPL* Triple negative	 15 (75) 2 (10) 2 (10) 1 (5)	 70 (77.8) 12 (13.3) 7 (7.8) 1 (1.1)	 0.655
JAK2 V617F allele burden ≥50%	11 (73.3)	47 (67.1)	0.64
HMR mutation	13 (65)	32 (35.6)	0.015
NDM NDM ≥3	20 (100) 12 (60)	64 (71.1) 9 (42.9)	0.006 0.0001

Table 3.3. *RAS*-mutated patient baseline clinical and mutation characteristics.

In patients with both molecular analysis and 24-week response/adverse event (AE) data available, the highest rates of ≥35% spleen volume reduction (SVR) were observed in the 200mg BID arm: 12.8% (n=5/39) followed by 2.8% (n=1/36) in 100mg BID arm and 0% (n=0/34) in 100mg QD arm). There were no significant correlations between driver or NDM mutation status and SVR or symptom responses, although numbers of events for analysis was low.

Grade 3/4 anaemia was reported in 13.6% (n=15/109) patients with molecular analysis; this more often occurred in *TET2*-mutated patients (odds ratio, OR, 4.2, 95% CI 1.4-13, *P*=.012), Figure 3.4B, an observation not previously been reported with other JAK inhibitor treatments[185].
Grade 3/4 thrombocytopenia was reported in 20% (n=22/110) and a haemorrhagic grade 3/4 adverse event occurred in 30.6% (33/108). Specifically, *RAS*-MT patients more often had a grade 3/4 thrombocytopenia (OR 3.64, 95% CI 1.2 – 11.3, *P*=.026), Figure 3.4B, and haemorrhagic AE (OR 3.59, 95% CI 1.28 – 10, *P*=.015). This association of RAS-MT retained significance when adjusted for baseline platelet count <50 $x10^9$/L which significantly associated with grade 3/4 thrombocytopenia (OR 4.7, 95% CI 1.7 – 13, *P*=0.002) or haemorrhagic AE (OR 2.5 95% CI 1.08 – 5.7, *P*=.033) on univariate analysis. Grade 3/4 infection AEs occurred in 15.5% of patients. No specific mutation subtype was associated with infection AE. However, ≥3 NDMs were associated with infection AE; OR 7.59, 95% CI 2.45 – 23.4, *P*=.0001. Due to limited length of follow-up, only 4 deaths were reported, precluding meaningful survival analysis.

Cytokine data was analysed using unsupervised learning identifying 6 clusters (Figure 3.4C). Among these, elevations in clusters 2 and 4 were associated with presence of HMR mutations. Higher cluster 2 scores were associated with driver mutation variant allele frequencies ≥50%. Notably, the pro-inflammatory cytokines in cluster 2 linked to HMR mutations (HMR+) represented a transcriptional cluster regulated by the NF-κB pathway. The presence of a HMR mutation was associated with higher IL-8 levels (40.5pg/ml) as compared with absence (24.5pg/ml), p<0.0001. Elevated tumour necrosis factor-alpha (TNF- α) and IL-18 levels were also associated with HMR mutations; TNF-α 61pg/ml in HMR+ vs. 48.5pg/ml for HMR-. Furthermore, cluster 2 scores which are enriched in cytokines involved in the NF-κB pathway were elevated in HMR+, p=0.02 (Figure 3.4C). Although RAS-pathway mutations were not associated with specific cluster scores, these patients did have higher levels of the NF-kB-associated cytokine IL12P40 (1.1ng/ml) as compared with RAS-pathway WT patients (0.6ng/ml), p=0.001. There was no association between cytokine cluster scores and recent exposure to RUX at trial entry.

Follow-up molecular analysis at week 24 was performed in 38.1% (n=42/110). No driver or NDM molecular responses were detected. Median change in driver VAF was 5% (range 0.25 – 73%). At least one new NDM was acquired in 7.1% (3/42), all mutations with a VAF of <5%; *CBL* [2],

TET2 [1], *TP53* [1], *U2AF1* Q157 [1]. No associations were observed between follow-up mutation analyses and efficacy outcomes or adverse outcomes.

3-2.5 *SF3B1* mutations in MPN

To examine the prevalence and specific clinico-genomic features of *SF3B1* mutations in MPN, I merged the MAJICET and PAC203 datasets with two other advanced MF trial cohort datasets (RESUME[182] and PRM151[186]). In the RESUME trial, red-cell transfusion dependent MF patients were randomised to pomalidomide vs placebo and were enriched for splicing factor mutations predominantly *U2AF1* (22%) followed by *SF3B1*. The PRM151 study was a phase 2 trial investigating the safety and efficacy of a recombinant human pentraxin-2 molecule as an anti-fibrotic therapy. This merged analysis consisted of 501 patients and *SF3B1* mutations were detected in 10.5% of patients (n=53); MAJIC-ET 6.3% (n=7/110), PAC203 12.7% (n=14/110), RESUME 10.7% (22/205) and PRM-151 13.1% (10/76) (Figure 3.5A). *SF3B1* K666 was the commonest hotspot mutation detected followed by K700 with other hotspots less frequently detected (Figure 3.5B).

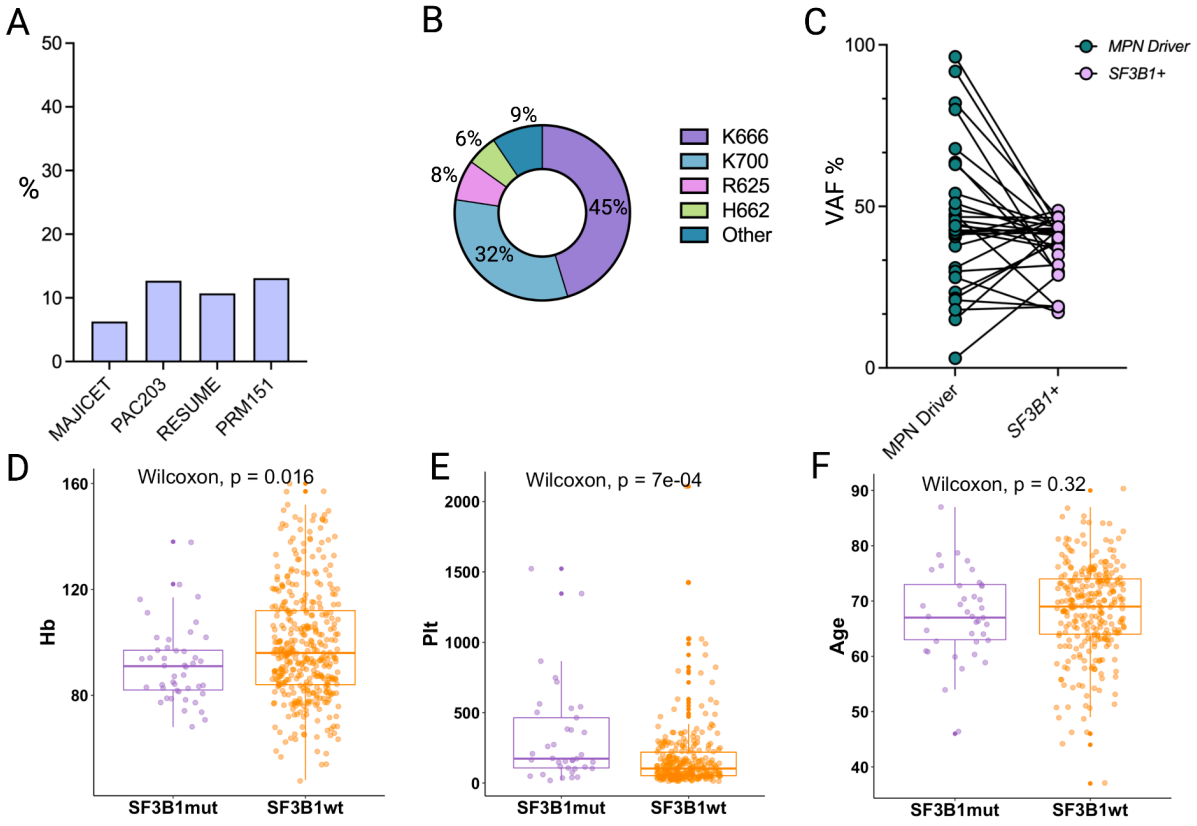


Figure 3.5. Clinico-genomic characteristics of *SF3B1* mutations in MPN. (A) Barchart showing the proportion of patients with *SF3B1* mutation by clinical trial; MAJIC-ET 6.3%, PAC203 12.7%, RESUME

10.7% and PRM-151 13.1%. (B) Donut chart showing the distribution of *SF3B1* hotspot mutations; K666 the most prevalent (45%) followed by K700 (32). (C) Dotplot illustrating the relationship between MPN driver VAF and *SF3B1* VAF; in 50% of cases, the MPN driver VAF > *SF3B1* VAF and in 50% of cases, the *SF3B1* VAF ≥ MPN driver VAF. Dot and boxplot of (D) median haemoglobin level in *SF3B1*mut (91g/L) vs *SF3B1*wt (96g/L) patients, p=0.016; (E) median platelet count *SF3B1*mut (214 x 10^9/L) vs *SF3B1*wt (164 x 10^9/L) MF trial patients and (F) median age in *SF3B1*mut (69, range 46 – 87 years) vs *SF3B1*wt (69, range 34 – 90 years), p=0.32.

MPN driver mutations were present in 94% of *SF3B1*mut cases; *JAK2*V617F in 58% (n=31), *CALR* in 17% (n=9), *MPL* in 5.6% (n=3) and 3.7% (n=2) were reported as TN. Mutation VAF was available for analysis in 295 patients. Comparing the MPN driver mutation VAF with the *SF3B1* VAF to infer probable order of mutation acquisition, I observed that the MPN driver mutation VAF was higher in 50% of cases and *SF3B1* was equal to or higher than the MPN driver in 50% (Figure 3.5C). Interestingly, when the MPN driver was higher, all cases were *JAK2*V617F but for the other group only 60% were *JAK2*V617F. When compared with *SF3B1*wt patients, *SF3B1*mut patients had lower haemoglobin levels (Figure 3.5D) and higher platelet counts (Figure 3.5E). There were no *SF3B1*mut age (Figure 3.5F) or gender associations.

3-3 Discussion

Distinct molecular signatures were identified in this analysis of treatment resistant MPN. In HC-RES/INT ET, this was the first description of the molecular landscape in context of a prospective randomized clinical trial, MAJIC-ET study[187]. NDM were higher than previous analyses likely relating to the particularly high-risk ET cohort enrolled who were exposed to multiple lines of prior therapy[43,163,165,188]. Importantly, I observed an enrichment for prognostically relevant mutations; *SF3B1* and *TP53* mutations, each detected at a frequency of 6.4% which is higher than reported previously in ET (prevalence of ~2 and 2-5% respectively)[43,188,189]. These mutations were strongly predictive of subsequent disease transformation highlighting the clinical/prognostic utility of mutation screening in HC RES/INT ET. Transformation risk was predominantly myelofibrotic risk as only one patient transformed to AML, although adequate assessment of leukaemia transformation events would require a longer follow-up period. *TP53* mutations in MPNs are well established as conferring a heightened risk of AML transformation with poor outcomes[43,112,188,190] but have not been associated with myelofibrotic transformation in ET[43,188]. SF mutations, most often *SF3B1,* have also been described by other groups to confer increased risk of myelofibrotic transformation[181,188,191]. A proposed prognostic scoring system was recently published[192]

combining mutation and clinical information to predict survival and myelofibrosis recognising the adverse impact of splicing factor mutations in ET and PV on overall (*SF3B1*, *SRSF2* in ET and *SRSF2* in PV) and myelofibrosis-free survival (*U2AF1* and *SF3B1* in ET) and the leukaemic risk associated with *TP53* in ET.

Crucially, transformation risk associated with these molecular profiles was not mitigated by RUX treatment in the study which is notable as there has been interest in the potential that early intervention with JAK2 inhibition might moderate the disease course. However, interestingly, *JAK2*V617F-mutated patients more often showed a platelet response as compared with non-*JAK2*V617F-mutated patients with RUX; the first association of driver mutation status with RUX platelet responses in ET. Moreover, non-*JAK2*V617F-mutated ET patients more often discontinued RUX treatment. In MF, RUX responses occur irrespective of driver mutation status[193]. These data suggest this differs for ET where *JAK2*V617F-mutated ET patients might specifically benefit from RUX.

Longitudinal mutational profiling in MAJIC-ET showed that most mutations, driver MPN and NDM, remained stable over time. Few driver MPN molecular responses were observed and loss of MR correlated with NDM clonal evolution. The follow-up period was insufficient to determine if this correlated with clinical disease evolution. In a more recent analysis of the PV arm of this study, MAJIC-PV[194], more patients had MPN driver molecular responses (>50% reduction in 56% of RUX treated patients) but NDM also influenced these molecular responses. The presence of NDM was associated with lower likelihood of developing a molecular response and a poorer event-free survival. In serial NDM analysis, 87.5% of “new” NDM were present at baseline below the pre-specified threshold for variant calling (5%). Considering these included mutations of prognostic value such as *SF3B1*, there may be clinical merit in lowering the diagnostic threshold for variant calling of certain prognostically relevant mutations.

The poor prognostic impact of *SF3B1* mutations in ET is discordant with the role of this mutation in myelodysplastic syndromes (MDS) where *SF3B1* mutations are associated with an indolent disease course and improved survival outcomes[195-200]. This implies that disease context and presence of additional mutations (primarily *JAK2* for MPNs) are important. *SF3B1* K700 is the most prevalent hotspot mutation in *SF3B1* MDS and a recent analysis observed that the milder phenotype was confined to patients with this hotspot[201]. Furthermore, a large analysis inclusive

of >1200 *SF3B1*mut MDS/AML[202], *SF3B1* K666 hotspot mutations were enriched in high-risk MDS and AML cases. Disease-specific hotspot *SF3B1* mutations are also described in non-haematological malignancies; *SF3B1* R625 mutations are predominant in uveal melanoma[203]. The hotspot-specific effect on disease type/phenotype may also be relevant in MPN in addition to co-occurrence with a JAK-STAT signalling mutation. In the analysis inclusive of 53 cases, I observed *SF3B1* K666 hotspot occurring most often followed by K700. A similar hotspot distribution was recently reported in a similarly sized retrospective analysis MF focusing on *SF3B1*[204]. In the MAJIC-ET dataset, K666 was the commonest hotspot present in 5/7 *SF3B1*mut cases but these numbers are too few to infer association with hotspot and risk of MF progression. For PAC203 MF patients, it was not possible to study hotspot specific associations with more advanced disease / survival due to a uniformly high-risk cohort and a short follow-up period.

The mutation landscape in a RUX treatment-resistant cohort, high-risk RUX RES/INT thrombocytopenia-enriched MF, showed high proportion of HMR mutations and an unexpectedly high frequency of *RAS/CBL*-pathway mutations (20.9%), which was higher than previously reported in MF (6 – 8.1%)[168,184]. HMR and *RAS/CBL*-pathway co-occurred and I observed for the first time, a mutation-specific cytokine signature in MPN; HMR mutations associated with a pro-inflammatory cytokine signature which likely reflects the inflammatory bone marrow micro-environment with pro-inflammatory signals derived from cells including the disease-driving clonal cells and also cells in the bone marrow niche[205]. It is possible that this inflammatory milieu plays a role in mediating RUX resistance.

Activating mutations of the *RAS/CBL*-pathway have been reported to correlate with reduced likelihood of spleen and symptom responses in patients with myelofibrosis treated with JAK2 inhibitors[184]. *RAS/CBL*-mutated patients particularly had a higher likelihood of developing grade 3/4 thrombocytopenia in the PAC203 cohort. *RAS/CBL*-pathway mutations in MF have been associated with shorter overall survival and higher likelihood of progression to leukaemia[168]. Together, these data support that *RAS/CBL*-pathway mutations typically occur as a late event in molecularly high-risk patients who are undergoing genetic clonal evolution. A potential underlying mechanism may be that JAK inhibition creates a selective pressure promoting evolution of *RAS/CBL*-pathway activating mutated subclones. This is supported by recent single-cell genetic analyses in myelofibrosis[206] showing enrichment in RAS mutations after JAK inhibition and in acute myeloid leukaemia (AML), *RAS/CBL*-pathway mutations emerge after treatment with FLT3-

inhibitors[207]. More recent data supports this hypothesis by studying the effect of JAK inhibition with RUX in cell lines and murine models revealing a fitness advantage gained in *RAS*-mutated clones after RUX exposure[208]. Therapeutic combinations targeting JAK and MEK/ERK in murine models recently have shown enhanced response in reduction of fibrosis, and this could be an effective strategy to mitigate risk of clonal evolution in high-risk patients[209].

In summary, the MAJIC-ET and PAC203 cohorts encompass molecularly high-risk patient group with distinctive molecular profiles. These findings will help inform the application of precision medicine for this challenging group of patients. Taken together, these data support the application of serial mutation analysis in HC-RES/INT ET patients as emergence of new, and prognostically informative NDM, occur even in patients with apparent driver MR.

Chapter 4: The molecular signatures associated with disease transformation in *JAK2*V617F-*SF3B1*-mutated MPN

4-1 Introduction

4-1.1 RNA splicing

RNA splicing is a process where introns, non-coding regions of the genome, are removed from precursor mRNAs to generate mature mRNAs which can be translated into proteins. Alternative splicing is where splicing occurs differentially in a single gene allowing generation of multiple proteins (Figure 4.1). This is an essential step in protein complexity across eukaryotes and important in multiple biological processes such as normal development, cellular diversity and cell apoptosis.[210]

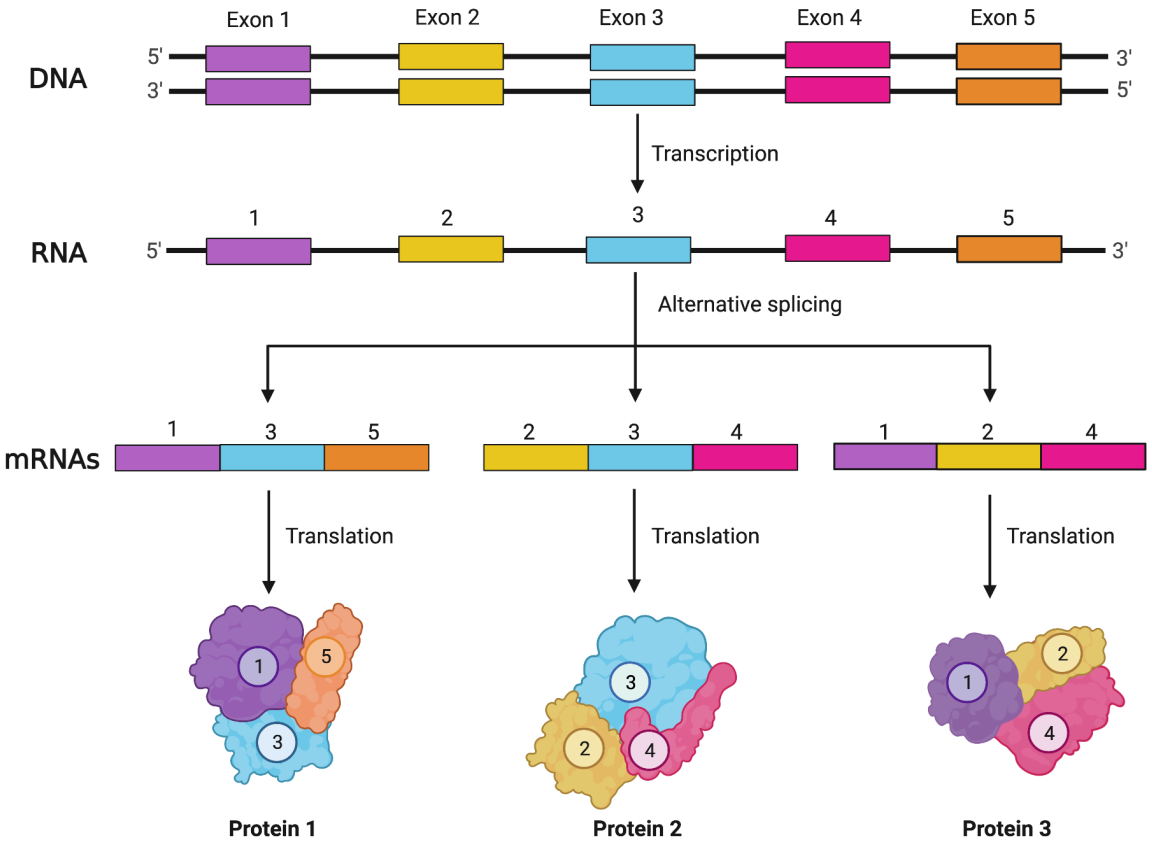


Figure 4.1. Schematic outline demonstrating alternative splicing of exons within single gene can generate multiple proteins. Image generated using BioRender.

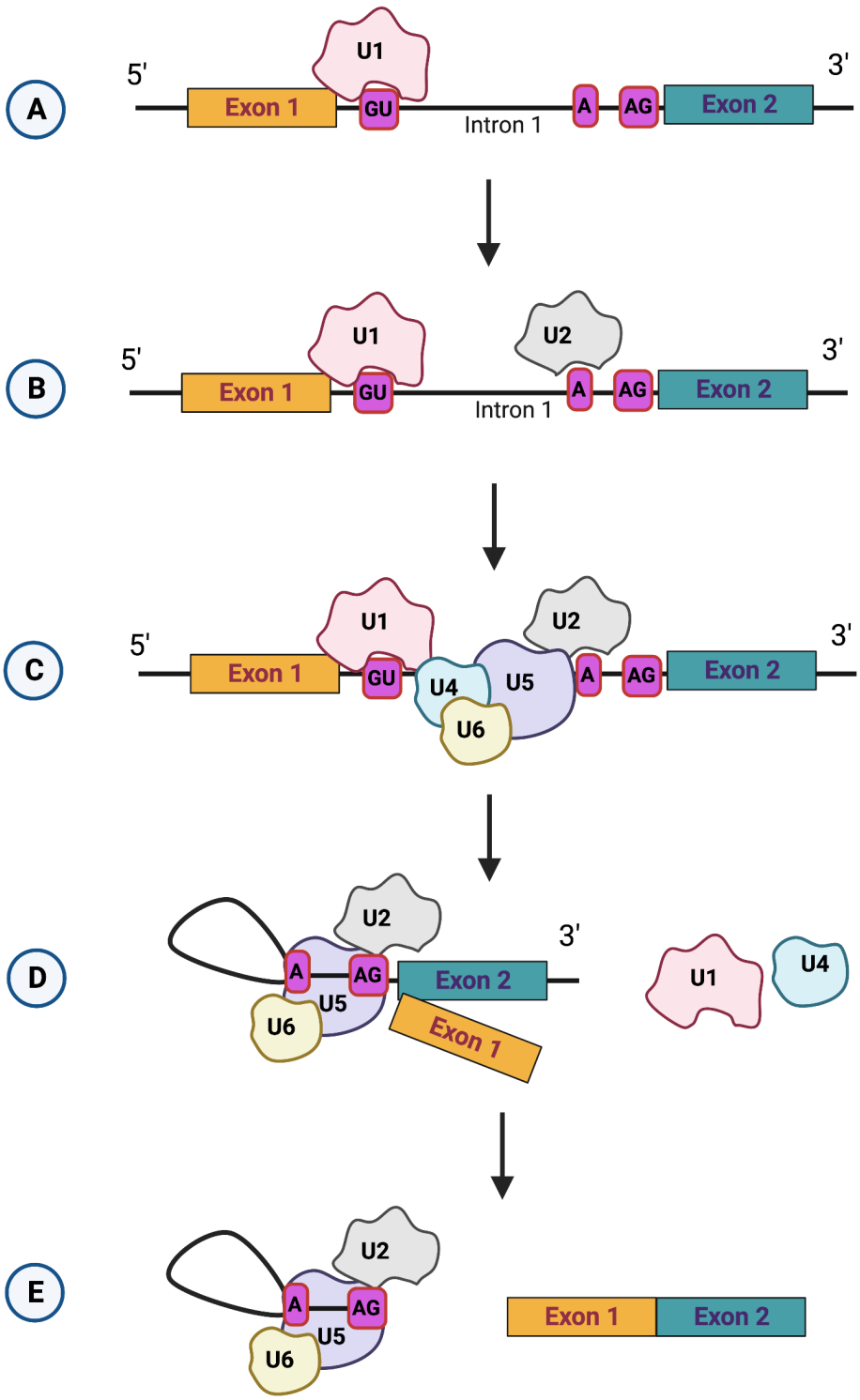


Figure 4.2. U2-dependent (major) spliceosome. (A) U1 snRNP recognises and binds to the 5' splice site consensus sequence (GU). (B) U2 snRNP next binds to the branch site sequence. (C) U5 and U4/U6 then associate with the complex with U1/U4 released (D) resulting in the final active spliceosome. (E) Intron removal then takes place followed by spliceosome disassembly. Image generated using BioRender.

Over 99% of splicing in humans is carried out by U2-dependent or major spliceosome, complex made up of five small nuclear ribonucleoproteins (snRNPs; U1, U2, U4, U5, U6) and multiple other non-snRNPs.[211] The minor U12-dependent spliceosome and accounts for <1% of splicing in humans. Briefly, the major spliceosome assembles on introns by sequentially recognising consensus sequences starting with U1 snRNP on 5' splice site (ss) at the 5' intron region followed by the U2 snRNP (containing SF3B1) binding to the branch site sequence[210] (Figure 4.2A, B). U5 and U4/U6 associate to be added to the complex and then U1/U4 are released to generate the final active spliceosome (Figure 4.2C, D). It is after this that intronic removal takes place, followed by disassembly of the spliceosome (Figure 4.2E).

4-1.2 RNA splicing is disrupted in cancer

RNA splicing is altered in cancer through multiple modes and next generation sequencing techniques have been instrumental in improving understanding of this process. In an analysis of over 30 tumour types in 8705 patients, alternative splicing was increased in tumours by up to 30% more than in normal tissues[212]. Furthermore, recurrent mutations have been identified in key components of the spliceosome in both solid and haematological malignancies[213]; most commonly in *U2AF1*, *SF3B1*, *ZRSR2* and *SRSF2* genes. These splicing factors (SF) have distinct mechanisms of action. However, when mutated they occur in an almost mutually exclusive fashion. *U2AF1, SF3B1, SRSF2* mutations are only detected in the heterozygous state. Murine models have shown that hemizygosity and homozygosity are lethal with WT allelic expression essential for survival[214,215]. These three mutations are primarily missense mutations occurring in hotspot regions suggesting gain of function effects. In contrast, *ZRSR2* mutations can be identified across the gene and more often are splice-site, frameshift insertions / deletions or nonsense mutations resulting in loss of function[216].

4-1.3 *SF3B1* is recurrently mutated across cancer

SF3B1 mutations are the most frequently mutated SF in cancers and in particular, haematological malignancies. *SF3B1* is mutated in approximately 30% of myelodysplasia (MDS)[217,218] a type of myeloid malignancy where there is reduced and abnormal blood cell production called dysplasia of one or more of the blood lineages leading to cytopenias and carries a risk of progression to

leukaemia. *SF3B1*-mutated MDS is associated with a specific clinical phenotype of anaemia, characteristic bone marrow morphology features of ring sideroblasts and an indolent disease course[219]. Consequently, SF3B1-mutated MDS has now been classified as a specific subtype of MDS[220].

Mutations occur in hotspot regions in the C-terminal HEAT domain. Interestingly, particular hotspots are enriched in certain cancers. In uveal melanomas, *SF3B1* R625 is the predominant *SF3B1* hotspot mutation[203,221] but K700E hotspot mutations are observed most often in MDS cases (Figure 4.3). The K666 hotspot is more prevalent in MPN as I have described in chapter 3 (45-50%) and far less prevalent in low-risk MDS at ~1.5%[202].

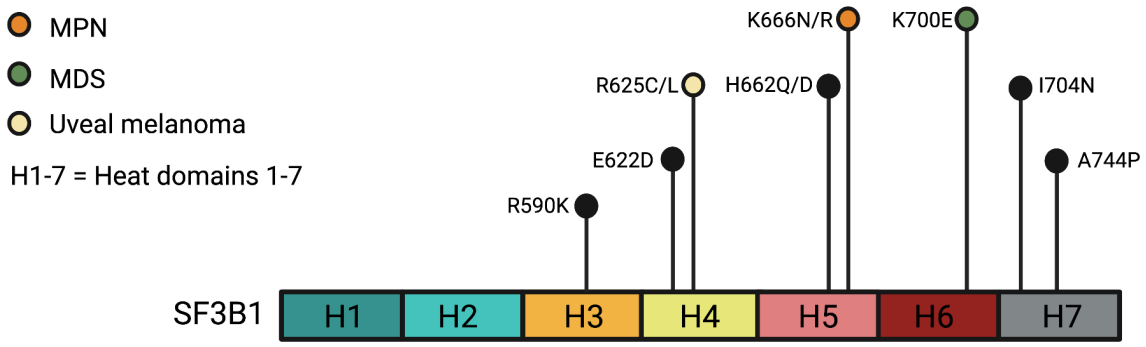


Figure 4.3. Lollipop diagram of SF3B1. Heat domains 1-7 with the most prevalent hotspot mutations shown. Hotspot mutations enriched in MPN, MDS and uveal melanoma highlighted.

In the mutated state, the most commonly abnormal impact of *SF3B1* on splicing is the use of a cryptic aberrant 3' splice (A3SS) 10-30 base pairs upstream of the canonical splice site[222]. In approximately 50% of transcripts, this causes introduction of a premature stop codon leading to a shortened transcript which is subject to nonsense-mediated decay (NMD)[223], a surveillance system to eliminate these transcripts[224]. Eventhough A3SS are the most prevalent *SF3B1*-mutant events described, other alternative splicing events can occur including downregulation in transcripts containing retained introns which have been identified in MDS[225] and CLL[226]. Aberrant splicing effects of *SF3B1* mutations will be explored in greater depth in chapter 5.

4-1.4 *SF3B1* mutations in myeloid neoplasms

In MDS, *SF3B1* mutations are disease-initiating and acquired within the HSC compartment[227] with evidence that a lymphomyeloid-biased HSC is the cell of origin[228]. In MPN, a stem cell origin for *JAK2*V617F and presence throughout the haematopoietic tree has been established[6,25,27]. *JAK2*V617F-*SF3B1*-mutated (JS+) clones have been detected in MPN HSPCs[122] but the cell of origin and the clonal hierarchy in MPN has yet to be well delineated. Understanding this is important since it may identify new cellular and molecular mechanisms of disease pathogenesis and novel therapeutic targets.

Deciphering the effect of *SF3B1* mutations on haematopoietic differentiation and the cellular composition of the haematopoietic compartment including progenitors and mature cells is essential to understanding the role of *SF3B1* acquisition in promoting MPN myelofibrotic transformation. Appreciation of disrupted haematopoeisis in *JAK2*V617F-mutated (J+) MPN has been advanced primarily due to single-cell genomics technologies which enable identification of the rare disease-propagating stem cells. Importantly, a bias towards megakaryocyte differentiation has been shown in both *JAK2*V617F- and *CALR*-mutated MF[124]. Expanded MF MkP populations have gene expression signatures associated with proliferation and fibrosis which is in line with prior data suggesting that MkPs are key drivers of fibrosis development in MPNs[229,230].

Further, single-cell genomics techniques may provide new insights into *SF3B1*mut-associated aberrant splicing within disease-driving stem cell populations and linking the effect of aberrant splicing with specific cellular populations or differentiation trajectories in the haematopoietic compartment. Uncovering mutant-specific aberrantly spliced transcripts may progress understanding of the disease biology and actionable therapeutic targets such as signalling pathway genes, activated oncogenes and neoantigens. The lattermost are especially relevant for *SF3B1*-mutated neoplasms. Neoantigens are proteins exclusively produced by cancer cells, not healthy cells, towards which immune responses can be triggered. Although the commonest source are somatic mutations, alternative splicing is another route. An aberrantly spliced transcript may produce a novel protein that is mutant cell specific and if immunogenic is classed as a neoantigen. *SF3B1*-specific neoantigens have been identified in MPN[231] but these studies were

performed at the 'bulk' level and not in individual cells. Identifying these targets in disease-driving rare stem cell populations is key to an effective therapeutic target.

Methods to interrogate splicing at the single-cell level are ever advancing and multi-omic approaches that study parallel mutation information and transcriptomics are particularly powerful for detection of mutant-specific events. These include methods such as the plate-based TARGET-seq approach and a new droplet method, GoT-splice[127], which integrates GoT[121] with long read sequencing[128]. I will explore the impact of *SF3B1* mutations on aberrant splicing in specific populations in more detail in chapter 5.

4-1.5 Hypothesis and aims

I hypothesise that *JAK2*V617F and *SF3B1* mutations collaborate to disrupt the HSPC compartment and promote myelofibrotic transformation, specifically:

1. Exacerbate *JAK2*V617F-mediated aberrant megakaryopoiesis
2. Exacerbate *JAK2*V617F-mediated aberrant erythropoiesis
3. Enhance downstream signalling in pro-fibrotic or inflammatory pathways

Aims:

1. To apply a multi-omic single-cell approach combining short-read transcriptomics and cell surface proteomics to describe the cellular landscape in JS+ MPN of stem and progenitor cells as well as mature lineage populations and their associated trajectories.
2. To study the associated gene expression signatures associated with JS+ cellular populations.

4-2 Results

4-2.1 Baseline characteristics of the dataset

To improve the power of the analysis to detect *SF3B1*mut MPN specific signatures, I extended the cohort for single-cell analysis was to any MPN subtype and confined the analysis to patients with *SF3B1* K666 hotspot mutations. I applied a high throughput droplet single-cell mutli-omic approach consisting of short read transcriptome sequencing (10X genomics chromium platform (version 3.1) paired with surface proteomics by sequencing CITE-seq[140] to a cohort of 12 MPN

patients inclusive of six JS+ MPN patients, six J+ MPN patients as control MPN cases and 5 healthy age-matched controls. Patients were matched for treatment, disease stage and tissue type (BM / PB). TARGET-seq was performed in parallel on these samples and the results are detailed in chapter 5.

SampleID	Mutation (VAF)	Donor	Tissue	Diagnosis	Treatment	Age	Gender	Single-cell approach
IF07008	-	HD	BM	-	-	65	M	10X + TS
IF09007	-	HD	PB	-	-	58	F	10X + TS
IF09015	-	HD	PB	-	-	56	F	10X + TS
IF09010	-	HD	PB	-	-	57	M	10X + TS
IF07003	-	HD	BM	-	-	79	F	10X + TS
TNO183	J+ (26%)	Pt	BM	ET	ANA	65	M	10X + TS
IF01132	J+ (48%)	Pt	PB	PPVMF	RUX	74	M	10X + TS
IF01109	J+ (33%)	Pt	BM	pPMF	pIFN	53	F	10X + TS
IF01089	J+ (54%)	Pt	PB	pPMF	IFN	63	M	10X + TS
TNO168	J+(42%)	Pt	BM	ET	ANA	68	M	10X + TS
TNO163	J+(65%)	Pt	BM	PETMF	RUX	73	F	10X
TNO123	J+(16%), S+ (17%)	Pt	BM	ET	ANA	77	M	10X + TS
IF01023	J+ (82%), S+ (41%)	Pt	PB	PMF	RUX	86	M	10X + TS
IF01144	J+ (34%), S+ (39%)	Pt	PB	pPMF	pIFN	66	M	10X + TS
MW	J+ (76%), S+ (43%)	Pt	PB	pPMF	None	69	M	10X + TS
TNO103	J+ (42%), S+ (41%)	Pt	BM	ET	RUX	69	F	10X + TS
TNO087	J+ (80%), S+ (43%)	Pt	BM	PETMF	RUX	79	M	10X

Table 4.1. Summary of patients and healthy controls used for 10X/CITE-seq and TARGET-seq experiments. Age reported from time of tissue sampling. VAF=variant allele frequency, J+ =*JAK2*V617F, S+=*SF3B1*, HD=healthy donor, Pt=patient, PB=peripheral blood, BM=bone marrow, ET=essential thrombocythaemia, PPVMF=post-polycythaemia myelofibrosis, pPMF=pre-fibrotic myelofibrosis, PMF=primary myelofibrosis, RUX=ruxolitinib, pIFN=pegylated interferon-2a, F=female, M=male, TS=Target-seq

Total MNCs and CD34+ cells were sorted in equal numbers; ranging 2,500 – 25,000 cells per population per donor. CD34+ antibody-derived tag (ADT) marker was included in the experiment with other progenitor markers to improve characterisation of the HSPC compartment. However, since the cells were stained with CD34 APC-efluor780 fluorophore to enable sorting by CD34+ cell surface protein expression, the signal of the fluorophore may have been affected by the ADT marker due to competitive binding at the same epitope sites. CD34+ ADT and fluorophore used recognised the same epitope (class III) but had different clones. Therefore, I tested the effect of staining cells with both markers on CD34+ APC-elfluor780 expression. There was a marginal,

non-significant reduction in staining of the positive population and I proceeded with a titration of 1:100 for CD34 APC-elfluro780 and 1:166 for CD34 ADT.

4-2.2 Quality control and donor demultiplexing

After sequencing, cell capture per library ranged 2,884 – 18,025 cells per library, in total sequencing 134,478 cells. After aligning raw fastq files to the genome, quality control (QC) and filtering was performed to remove low quality cells and doublets. Doublets are an important confounder to exclude in single cell transcriptomics. These are typically classed into homotypic, where the cells are transcriptionally alike (either from the same donor or cell type), or heterotypic doublets where the cells are transcriptionally distinct (either distinct cell types or different donors)[145]. Doublet detected methodologies can be experimental and computational.

I incorporated two experimental methodologies into my experimental design; Souporcell[146] which leverages the donor genetic SNP information from the RNA-sequencing reads allowing for cross-donor doublet detection and donor demultiplexing downstream and a second called Cell Hashing[144] which involved labelling the cells with oligo-tagged antibodies. In the latter method, the antibodies are reactive to proteins expressed universally in all cells and therefore the hashing label can uniquely stain the cells of each donor facilitating pooling of different donors / cell types for sequencing with downstream demultiplexing by detection of the unique oligonucleotide tag / barcode. Both experimental and computational methods are more sensitive for heterotypic than homotypic doublet detection.

For final sample demultiplexing and doublet exclusion, I used Souporcell as this method reliably assign all cells correctly to each donor. It outperformed Cell Hashing for cell assignment with an average proportion of unassigned cells of 10.3% for hashing (ranging 1.7 - 33.3%) as compared with an average 0.75% for Souporcell (ranging 0 – 2.1%) (Figure 4.4A,B). Plotting gender-specific genes in an example library containing 2 donors demonstrates Souporcell faithfully assigns cells correctly with fewer unassigned cells than Cell Hashing (Figure 4.4C). The mean doublet rate was 10.4% (range by library 1.4 – 28%) for Souporcell and 7.5% (range by library 2.4 – 13.8%) for Cell Hashing. I compared these methods with a computational approach, DoubletFinder[232], which is based on the principle of finding gene expression similarities. The mean doublet rate was similar using this approach at 11.1% (range by library 3.9 – 24%).

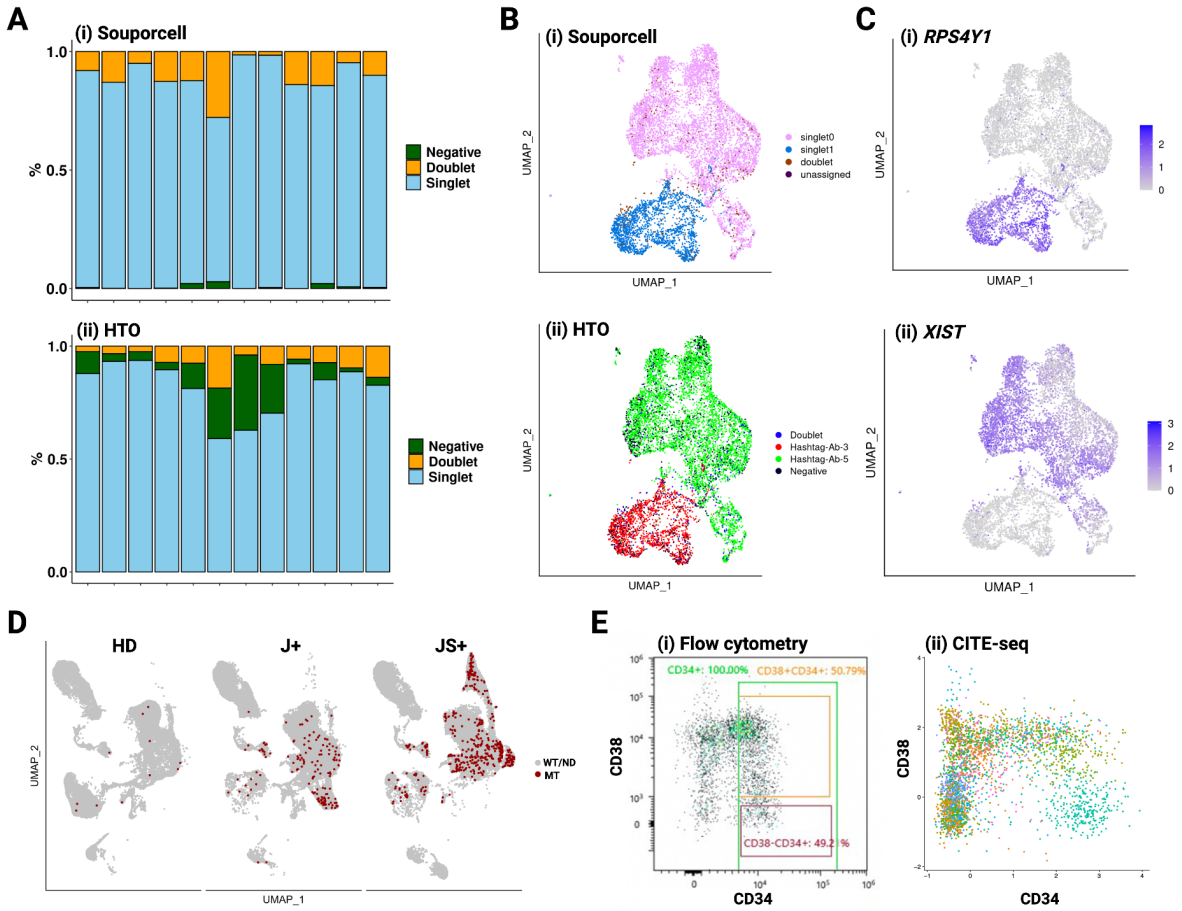


Figure 4.4. 10X data donor demultiplexing and quality control. (A) Bar chart showing the proportion of singlet, doublets or negative cells as determined by (i) Souporcell or (ii) HTO. (B) UMAP plots of cells from a library containing two donors; cells are coloured by donor using (i) Souporcell and (ii) HTO. (C) The same UMAP plot showing gender genes; the male gene (i) *RPS4Y1* correctly highlighting male donor cells and (ii) *XIST* correctly highlighting female donor cells. (D) UMAP plot highlighting *JAK2*V617F mutation by donor in 0.03%, 0.43% and 1.1% of HD, J+ and JS+ cells respectively, p<0.0001. (E) Dot plots of CD34+ expression (x-axis) and CD38+ expression (y-axis) comparing (i) flow cytometry expression of CD34+ primary cells from an individual donor with (ii) CITE-seq expression of merged CD34+ and total MNCs from a different individual donor.

QC filtering was carried out by applying the following thresholds to remove doublets, low quality and apoptotic cells: singlets by Souporcell, genes ≥ 500, UMI ≥1000 and MT <15%. Applying these filtering criteria, 109,498 cells passed QC for downstream analysis.

Prior to downstream gene expression and CITE-seq analysis, I used a variant calling tool to ascertain if I could detect reads at the *JAK2* and *SF3B1* allele position and therefore, the mutation.

However, detection of SNV mutations from high throughput scRNA-sequencing has low sensitivity[233] without addition or mutation-specific primers to enrich the region of interest. Using vartrix, a single-cell variant calling software (https://github.com/10XGenomics/vartrix), I was able to detect reads at the *JAK2* position in 1.8% of cells passing QC (2.4% in HD cells, 1.3% in J+ patient cells and 1.8% of JS+ patient cells). *JAK2*V617F mutation was present at a frequency of 0.03% (n=11/35331 cells), 0.43% (n=152/34115 cells) and 1.1% (n=462/40052 cells) of HD, J+ and JS+ donors respectively , p<0.0001 (Figure 4.4D). *JAK2*V617F-mutant reads in HD cells was unexpected but this was confined to donor IF07003 who also had the mutation detected in the matched TARGET-seq data providing orthogonal evidence this is likely a case of CH. The observation that more *JAK2*V617F-mutant cells were detected in JS+ donors will be corroborated with TARGET-seq data described in chapter 5 but one hypothesis may be that cells carrying a prior *SF3B1* mutation are more likely (or "primed") to acquire a *JAK2*V617F mutation. There was sufficient coverage at the *SF3B1* positions for K666 in 2.1% of cells and in 12% of these cells (n=271/2250), an *SF3B1* mutation was detected.

4-2.3 CITE-seq

To improve characterisation of cellular populations and expression of novel immune checkpoint markers, I included 46 CITE-seq markers (Chapter 2, Table 2.1). Determining optimal titration for each ADT marker in a manner analogous to flow cytometry was not feasible. Therefore, I included each ADT marker at 30% of the recommended concentration by Biolegend based on an analysis on 52 markers establishing that reducing the concentration to 6.25 - 25% of Biolegend recommended improved signal-to-noise by significantly reducing background staining without compromising antibody sensitivity[141]. I aimed to sequence the libraries to the recommended depth of ~5000 reads per cell[125,140]. Overall, there was mean and median of 6,974 and 6,206 reads per cell respectively with a range per library of 1,711 – 26,932 reads per cell). CITE-seq markers mirrored canonical HSPC flow cytometry marker expression illustrated by example of a sample where I plotted CD34 and CD38 expression showing expression by flow cytometry and CITE-seq for the same sample (Figure 4.4E).

4-2.4 Data integration

Integrating samples in single cell datasets is key to study and compare the cellular populations and gene expression signatures between mutated cases and healthy controls. An important confounder which may cause technical rather than biological variation is batch effect which occurs through processing multiple samples and at different times. Relevant batch variations can occur from initial sample processing and library preparation to sequencing steps. Prior to proceeding with integration, I merged the libraries without batch correction (Figure 4.5A).

I compared two RNA integration methods on the dataset to apply for control of batch effect; reciprocal principal component analysis (RPCA)[147] (Figure 4.5B) and harmony integration[148] (Figure 4.5C). RPCA identifies shared features “anchors” between datasets and projects features onto a shared PCA space using mutual nearest neighbour principles. Harmony embeds cells from the datasets into a shared two-dimension space and assigns cells to multiple clusters. Within each cluster, a correction factor is calculated and cells are iteratively reassigned until convergence is achieved. Harmony most faithfully integrated the samples together to achieve the clearest resolution for each cell population, determined using visual inspection and calculating cLISI and iLISI scores (Figure 4.5D). Harmony had cLISI score closest to 1 and the highest iLISI score indicating better integration then RPCA. Visual inspection included overlay of lineage signature genes on UMAP plots (Figure 4.5E), a dot plot (Figure 4.5F) and ADT markers (Figure 4.5G) on UMAP plots.

To establish whether incorporating both RNA and ADT information improved cell clustering, I performed multimodal integration using both RNA and ADT information (Figure 4.5H). Weighted nearest neighbour analysis (WNN) from Seurat[125] applies an unsupervised approach to weigh RNA and ADT information separately, determine the relative weights per cell and integrate the information on a shared WNN graph. WNN clustering was not able to distinguish specific HSPC populations such as EBM (not detected in “HSPC” cluster as highlighted in Figure 4.5I) when the same cell clustering resolution (0.8) was applied as for RNA (EBM and HSPC clustered separated as highlighted in Figure 4.5J) suggesting it may be inferior for these progenitor populations in my dataset. However, it did show potentially enhanced resolution in mature CD4+ T cell, CD8+ T cell and NK cell populations. For downstream analyses on HSPC data, I used harmony RNA integrated cell clustering and applied WNN clustering for mature cell population analysis.

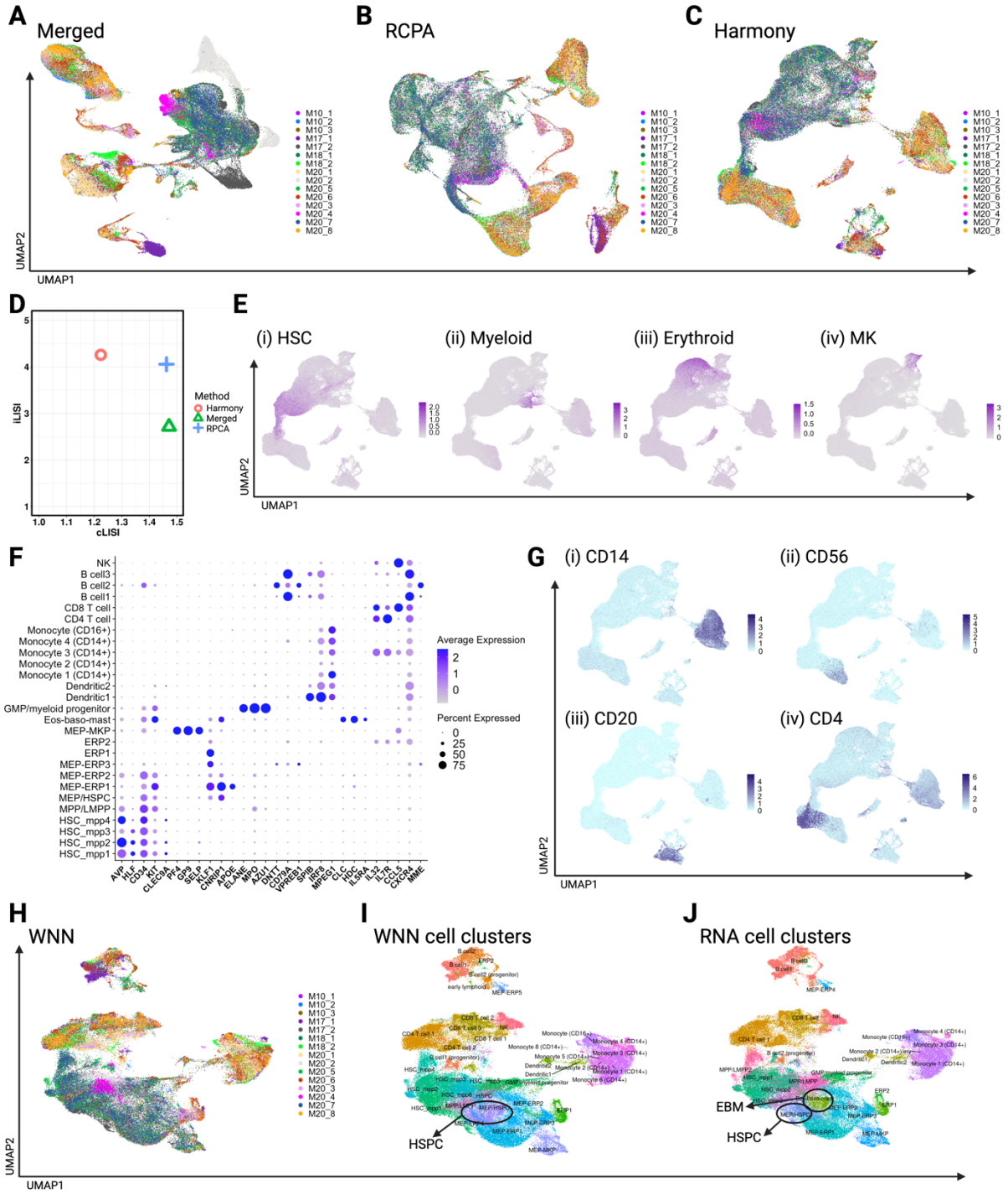


Figure 4.5. Data integration and cell cluster annotation. UMAP plot of cells coloured by library batch merged (A) without integration, (B) integrated with RPCA and (C) integrated with Harmony. (D) LISI analysis on the 3 methods, cLISI score and iLISI scores shown on x- and y-axis respectively. An optimal integration method should have a cLISI score close to 1 and a higher iLISI score. (E) Harmony UMAP plot of selected lineage signature gene expression: (i) HSC, (ii) Myeloid, (iii) Erythroid and (iv) MK. (F) Dot plot of canonical lineage marker gene expression for cell type. The dot sizes represent the percent of cells expressing a

gene. (G) Harmony UMAP plot of select mature lineage ADT marker expression. (H) UMAP plot of cells integrated with WNN and harmony methods. Harmony WNN UMAP plots of WNN analysis derived cell clusters (I) and of harmony RNA derived cell clusters (J) highlighting specific cell clusters.

4-2.5 Characterising HSPC and mature lineage compartments

After integration and dimensionality reduction, I applied graph-based clustering using louvain algorithm finding 30 discrete clusters. To annotate each cluster, I examined the marker genes per cluster and correlated these with lineage marker genes (derived previously[124]) and ADT markers to manually curate the annotation (Fig 4.5E-G).

Expected cell populations within HSPC and mature lineage compartments were present in both patients and healthy control cells (Figure 4.6A). In JS+ and J+ cells, specific HSPC subpopulations were expanded relative to HD cells, namely megakaryocyte-erythroid progenitor (MEP) and erythroid/megakaryocytic populations (Figure 4.6B). There was a particular expansion of JS+ ErP, MkP and EBM progenitor cells relative to J+ and HD cells (Figure 4.6B). I determined the AUCell score of lineage signatures for each cell. This score is derived by ranking the expression of each genes a gene-set for each cell and is independent of clustering or normalisation methods. There was a significant increase in JS+ cell AUCell ErP scores (9.3%) relative to J+ (6.8%) and HD cells (6.2%), $p<0.0001$ (Figure 4.6B). AUCell MkP scores showed non-significant increase in JS+ cells (3%) relative to J+ cells (2.6%) but both were increased relative to HD cells (0.7%), $p=0.001$. I observed the most striking AUCell score increase in EBM progenitors with an over 3-fold increase in JS+ as compared with J+ or HD cells, $p<0.001$ (Figure 4.6C).

Using MiloR, I performed differential abundance (DA) analysis, a method which doesn't rely on discrete cell clustering but instead categorises cells into overlapping neighbourhoods embedded on a KNN graph. This is especially useful for interrogation of HSPCs populations which differentiate along a continuum rather than distinct stages and therefore detecting cell types/states associated with diseases which otherwise be missed by assigning HSPCs to discrete clusters. Moreover, MiloR accounts for technical variables such as batch effect and can control for bias introduced by sample-specific variation in cell numbers. DA analysis replicated the changes shown in cell clustering analyses demonstrating JS+ cells enrich in megakaryocytic, earlier erythroid and mast cell progenitors as compared with those of J+ or HD cells (Figure 4.6D).

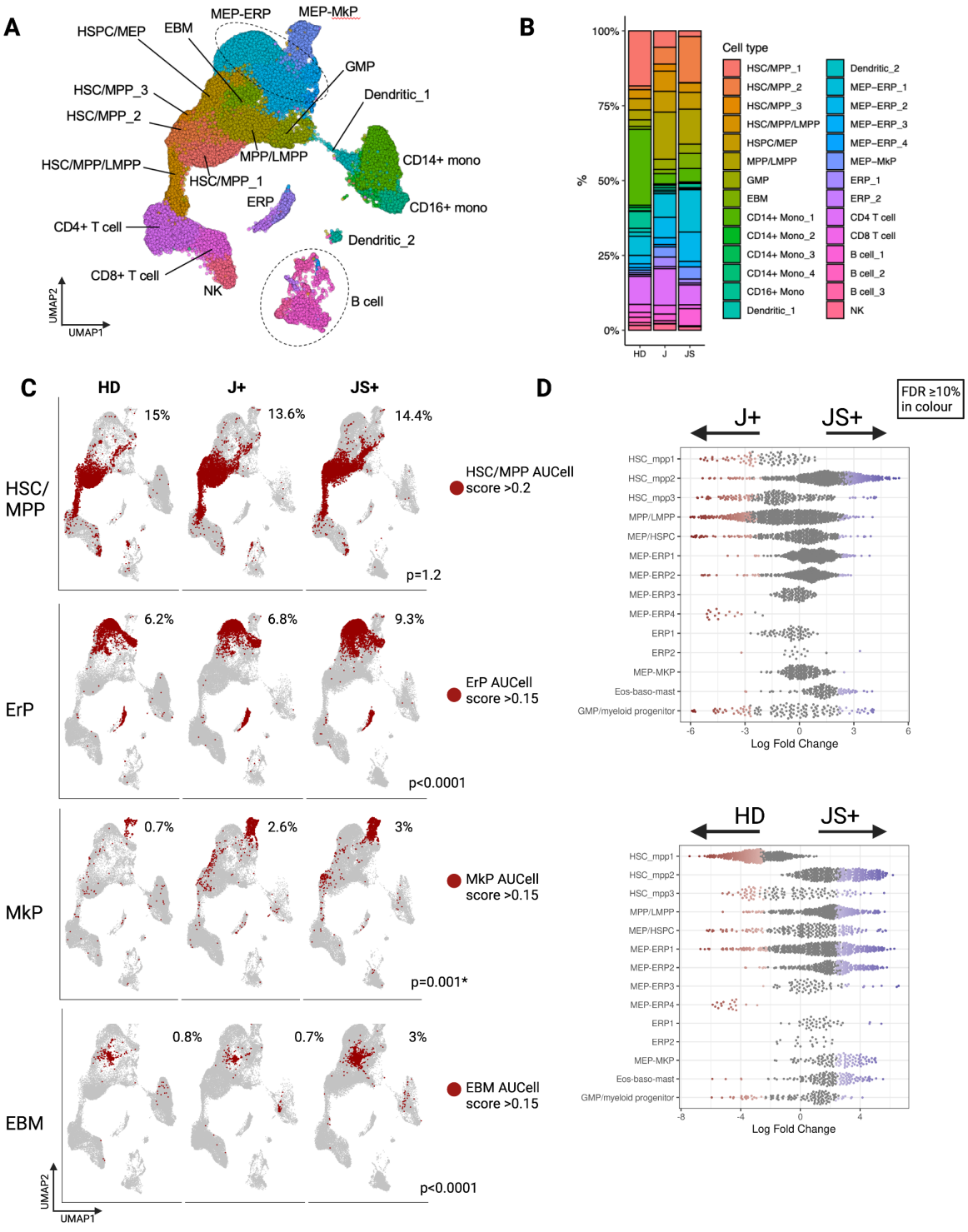


Figure 4.6. Characterisation of HSPC populations in the 10X dataset. (A) UMAP plot showing all cell populations. (B) Barchart showing frequency of cell populations by donor type. (C) UMAP plots by donor overlaid with HSPC AUCell scores. *=significance for JS/J as compared with HD cells only. (D) Beeswarm plots of differential abundance (DA) analysis results showing the log fold change across neighbourhoods containing cells from cell clusters. DA comparison between JS+ and J+ (upper plot) / JS+ and HD (lower

plot) performed with neighbourhoods at FDR 10% coloured purple if enriched in JS+ and red if enriched in J+ or HD.

Although HSC/MPP AUCell scores showed similar proportion of cells expressing HSC/MPP across each donor group (Figure 4.6B); the earliest HSC/MPP clusters were differentially enriched for either HD cells or JS+ cells (Figure 4.6B, D); 'HSC_mpp1' and 'HSC_mpp2' clusters were enriched for HD or JS+ respectively with equal numbers of J+ cells were in both clusters. Monocytes were reduced across both JS+ and J+ groups as compared with healthy controls and in the T cell compartment, CD4/8 T cells were overall increased in MPN samples as compared with controls (Figure 4.6B).

To determine lineage differentiation trajectories and in particular, whether JS+ MPN is associated with aberrant or expanded MkP/ErP/EBM lineage differentiation, I performed trajectory analysis on HSPC populations using force-directed graph (FDG). Cells from all samples were ordered from earliest HSPCs as they differentiate to mature cell populations. (Figure 4.7A).

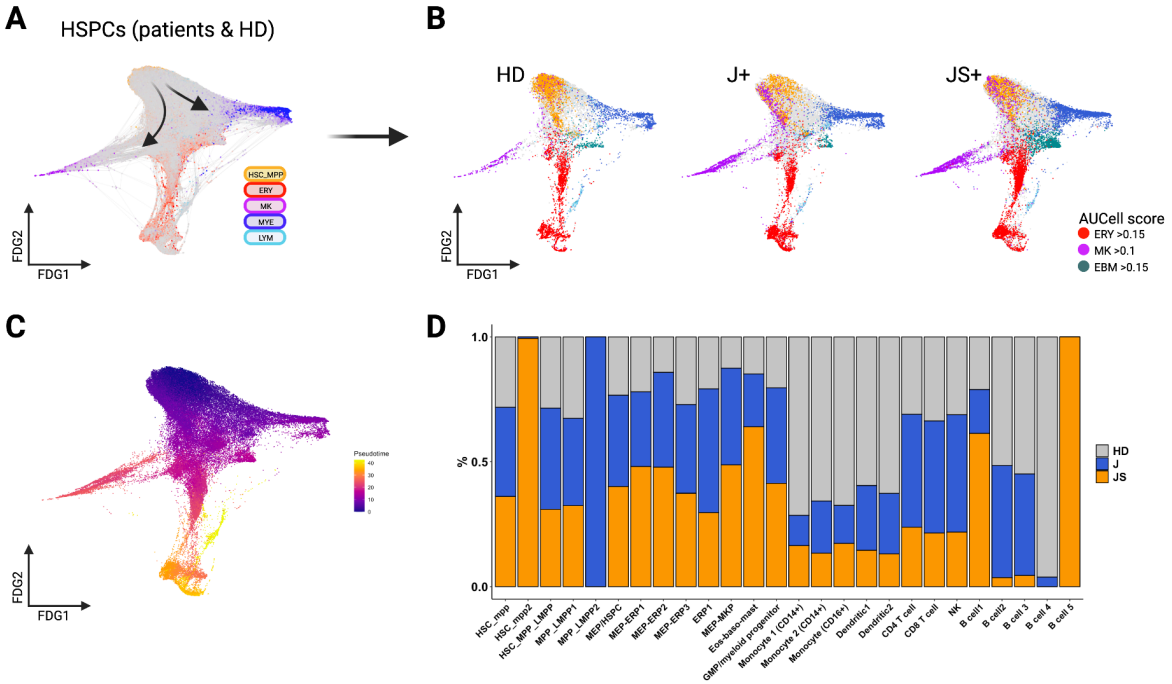


Figure 4.7. Lineage differentiation trajectories in 10X HSPCs. (A) Force-directed graph (FDG) of HSPCs with lineage genesets highlighted. (B) FDG by donor type showing expanded ErP and EBM in JS+ HSPCs. (C) Monocle pseudotime plotted on FDG coordinates of HSPCs. (D) Barchart of down-sampled cells (20,000 in each donor group) showing proportion of each donor per cell population.

Overlaying lineage signature scores showed the main lineage trajectories were from HSC/MPP to MK, erythroid and myeloid. JS+ cells were biased towards particularly expanded earlier erythroid and mast cell differentiation trajectories as compared with J+ or HD cells (Figure 4.7B).

I applied a second trajectory analysis method, Monocle3 pseudotime analysis, demonstrating the same trajectories across the dataset superimposing the pseudotime results of cell ordering on the FDG graph coordinates (Figure 4.7C). Finally, I down-sampled HD, J+ and JS+ total cells to equal numbers in each group (20,000 each), Figure 4.7D, and also to 10,000 for HSPC specifically (not shown) to remove the effect of cell number in JS+ HSPCs finding that the JS+-specific HSPC expansions were retained. This novel observation of expanded EBM progenitors in JS+ MPN may be relevant to the mechanism of disease progression since mast cells have been implicated in fibrosis development in a number of benign and malignant diseases[234,235,236,237,238,239].

4-2.6 Dysregulated gene signatures in JS+ HSPCs

To examine specific signatures associated with the aberrantly expanded JS+ HSPC populations, I performed differential gene expression analysis of HSPC populations comparing JS+ with *JAK2V617F* (J+) and healthy donor (HD) HSPCs and specifically focusing on molecular signatures distinct in ErP, MkP and EBM progenitors. Prior to performing specific progenitor analyses. I initially performed a HSPC analysis merging the progenitor subpopulations (n=62,121 cells) to enhance the power to detect differentially expressed signatures. For this HSPC analysis, 517 genes were differentially expressed specifically in JS+ HSPCs as compared with J+ or HD HSPCs (Figure 4.8A) with distinct molecular signatures in each donor group (Figure 4.8B). Global HSPC differential pathway analysis showed a downregulation in a number of pathways in JS+ HSPCs as compared with J+ HSPCs which may reflect cellular heterogeneity (Figure 4.8C); closer inspection of specific HSPC subpopulations showed this difference was confined to HSC/MPP populations only (Figure 4.8D).

Examination of JS+ HSPC-specific molecular signatures highlighted tryptase genes as some of the most upregulated genes (Figure 4.8E), being especially prominent in ErP and EBM populations. Tryptases are a group of serine proteases primarily released by mast cells. *TPSB2*, coding for the main isoenzyme beta tryptase and expressed in mast cells, was more highly

expressed than *TPSAB1*, a gene coding for both alpha and beta tryptase. Alpha tryptases are the major isoenzymes in basophils.

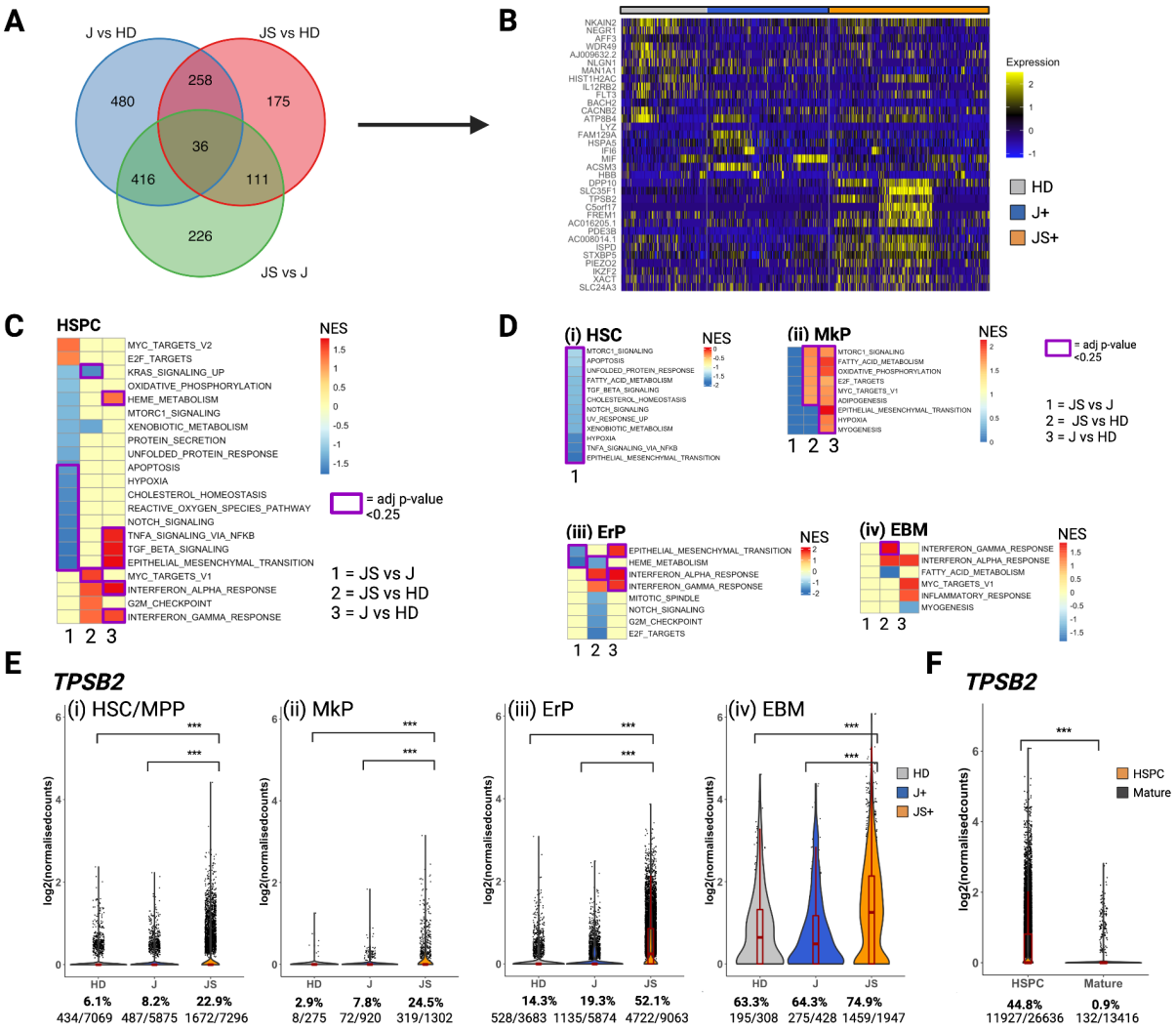


Figure 4.8. **Differential gene expression analysis in 10X HSPCs.** (A) Venn diagram of HSPC differentially expressed genes (DGE) between donors. (B) Heatmap of top DGE. (C) GSEA analysis showing downregulation in JS+ as compared with J+ HSPCs of multiple Hallmark pathways including apoptosis, TNFA and TGFB signalling. (D) GSEA analysis of HSC, MKP, ErP and EBM populations. Violin plots showing *TPSB2* enhanced expression was HSPC-specific (E, F); each dot represents the gene expression for each cell, boxplot in red represents median and quartiles. Percentage of cells expressing the gene presented below each plot. Wilcoxon test was used test statistical significance. NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

To confirm that increased *TPSB2* expression was confined to JS+ HSPC compartment, I compared expression of HSPC with mature populations showing that 44.8% of JS+ HSPCs

expressed *TPSB2* in contrast to 0.9% of JS+ mature cells ($p<0.0001$), Figure 4.8F. Mast cell tryptase has been implicated in fibrosis development in both benign[240,241] and malignant conditions[242]. Pertinent to MPNs, included in the latter group is systemic mastocytosis where malignant *KIT*-mutated mast cells are the disease-driving cells[243]. Patients often develop bone marrow fibrosis particularly in proximity to mast cells. Basophils have also been described to promote fibrosis development, particularly important in renal fibrosis[244] but also in pulmonary fibrosis and post cardiac transplant.

FREM1 gene was also one of the most highly upregulated genes in JS+ HSPCs (Figure 4.9A). This gene codes for an extracellular matrix protein and in normal developmental biology is involved in epidermal adhesion[245]. Whether it plays a role in normal or pathological haematopoiesis has yet to be determined but gene expression profiling of bulk RNA from BM MNCs of PMF patients (n=15) showed an increase in *FREM1* expression when correlating MF patient gene expression with lung fibrosis signatures[246]. Another relevant dysregulated gene in JS+ HSPCs which may be significant to exploration of mechanisms of bone marrow fibrosis development is *CLEC11A* which was upregulated in JS+ HSPCs as compared with J+ or HD HSPCs. (Figure 4.9B). This gene encodes for a C-type lectin named osteolectin, also called stem cell growth factor[247]. In murine models, it is expressed in myeloid cells and fibroblasts and supported the growth of haematopoietic progenitors in short-term liquid culture experiments[248]. *CLEC11A* plays a key role in osteogenesis with its requirement for murine mesenchymal stromal cell differentiation to osteoblasts[249]. It binds to integrin α11 (in humans and mice) which is expressed on leptin receptor positive stromal cells and osteoblasts[250]. Its expression was increased in acute myeloid leukaemia cohorts[251] and associated with an improved outcome.

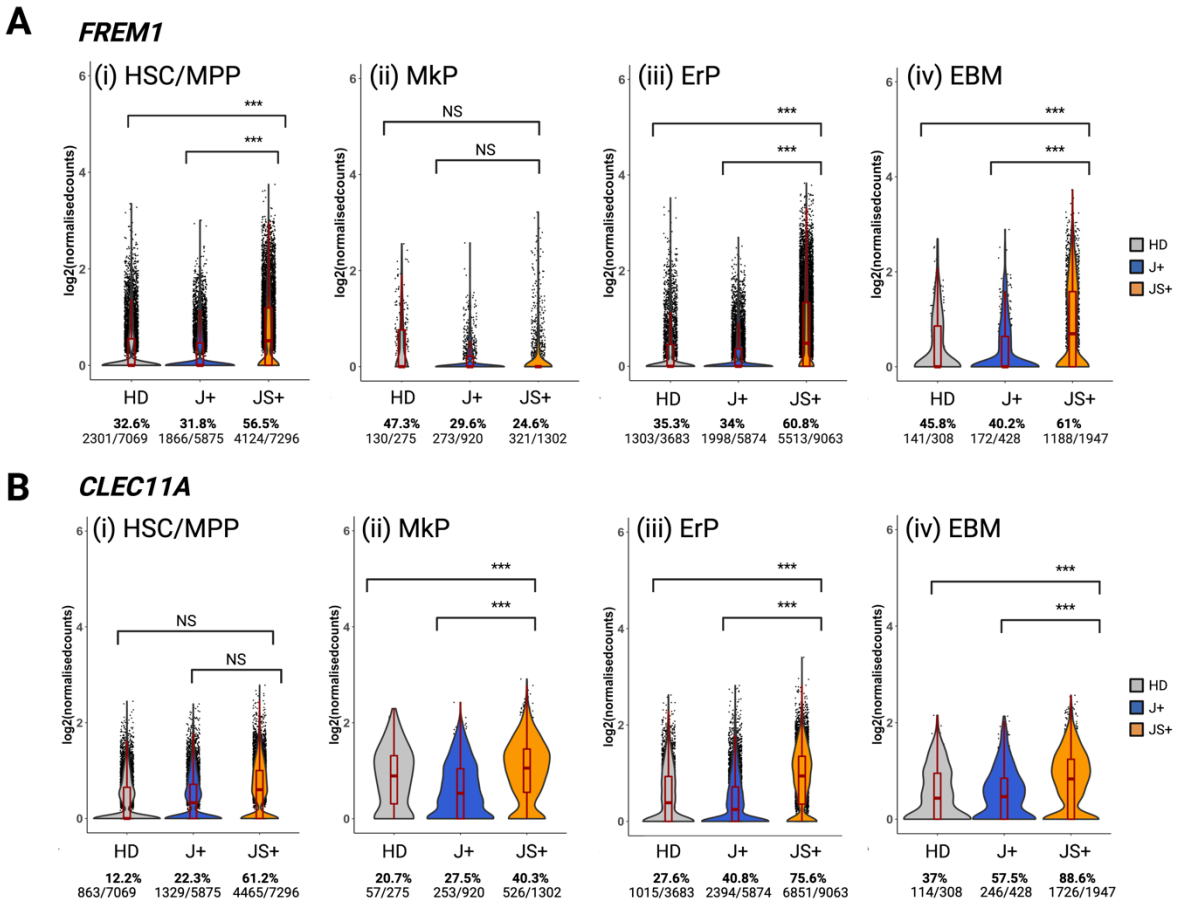


Figure 4.9. Gene expression of specific genes in 10X HSPCs. Violin plots showing (A) *FREM1* and (B) *CLEC11A* expression in HSC, ErP, MkP and EBM populations separated by donor. Each dot represents the gene expression for each cell, boxplot in red represents median and quartiles. Percentage of cells expressing the gene presented below each plot. Wilcoxon test was used test statistical significance. NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

Considering that the early HSPC clusters, HSC_mpp1 and HSC_mpp2, were differentially enriched for HD and JS+ cells respectively, I performed differential gene expression analysis between these two clusters, highlighted on the UMAP in Figure 4.10A. Although the clusters were enriched differentially for HD or JS+, this wasn't attributed to by a donor effect (Figure 4.10B). The proportion of J+ cells in each cluster was comparable in HSC_mpp1 and HSC_mpp2 clusters at 21.2% and 21.4% respectively.

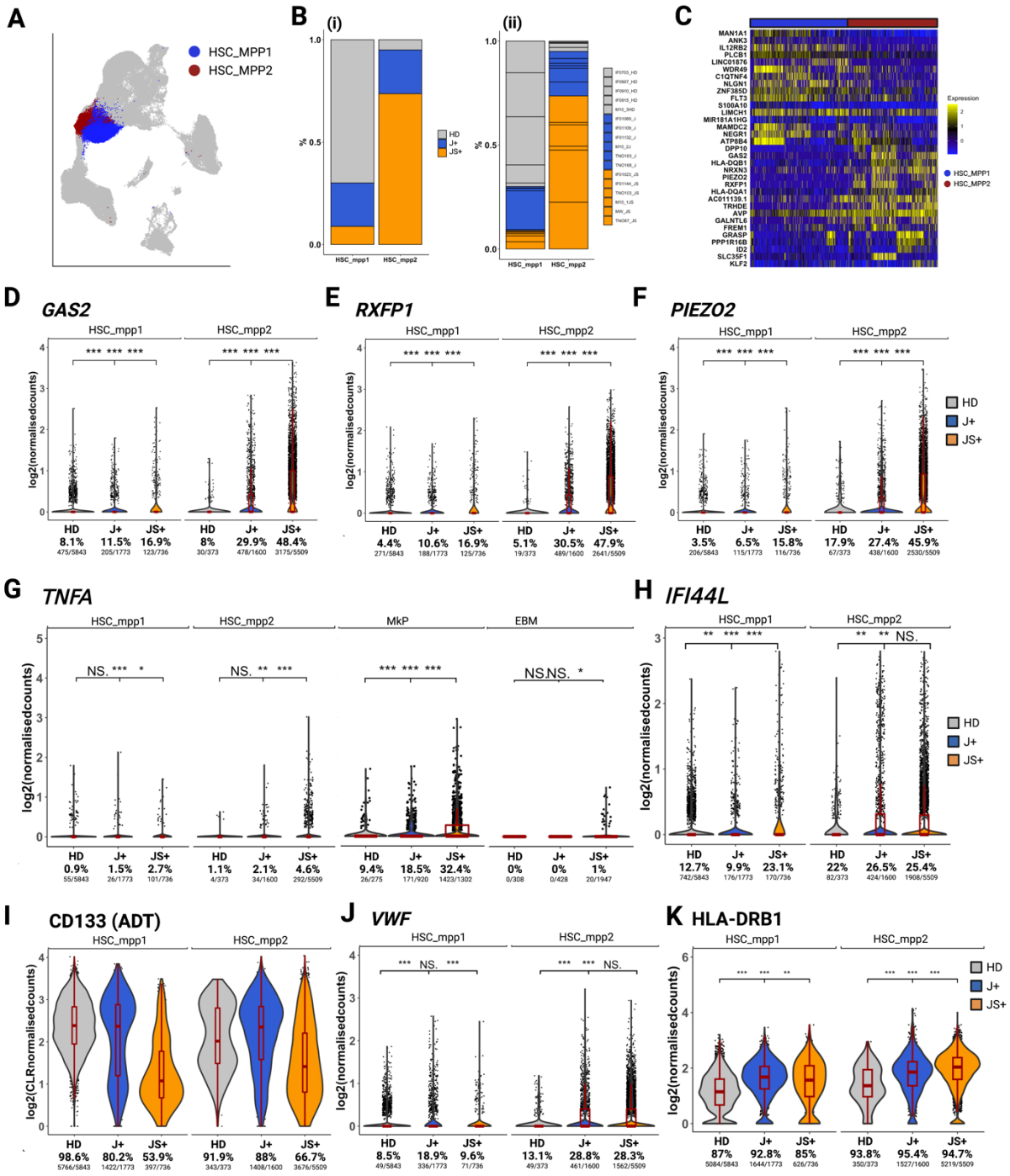


Figure 4.10. Differential gene expression analysis in 10X HSC/MPP. (A) UMAP plot highlighting HSC_mpp1 (blue) and HSC_mpp2 (red) clusters. (B) Bar chart showing frequency of overall donor genotype group (i) and each donor (ii) in HSC_mpp1 and 2 clusters. (C) Heatmap of top DGE between HSC_mpp1 and 2 clusters. (D) Violin plots showing top DGE genes in each cluster separated by donor group: *GAS2* (D), *RXFP1* (E) and *PIEZO2* (F). Violin plots showing *TNFA* expression in HSC_mpp1, HSC_mpp2, MkP and EBM populations separated by donor. Violin plots for each HSC_mpp cluster separated by donor group: *IFI44L* (H), CD133 ADT (I), VWF (J) and HLA-DRB1 ADT (K). Each dot represents the gene expression for each cell, boxplot in red represents median and quartiles. Percentage

of cells expressing the gene presented below each plot. Wilcoxon test was used test statistical significance, comparison displayed left to right (HD vs. J+, HD vs. JS+ and J+ vs. JS+). NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

Both clusters had distinct molecular profiles as expected considering the different contribution of healthy and patient cells per cluster (Figure 4.10C). *GAS2*, *RXFP1* and *PIEZO2* (Figure 4.10D-F) were among the most differentially expressed genes between these clusters and were most highly expressed in JS+ cells in HSC_mpp2 but also increased in J+ cells relative to HD cells in this cluster. These 3 genes were collectively upregulated in a bulk RNA-sequencing analysis of a related myeloid neoplasm, chronic myeloid leukaemia (CML)[252] and *GAS2* and *RXFP1* in single-cell analyses of haematopoietic stem cells of MPN[122] and CML[253] patients. These molecular signatures have been observed in BCR-ABL+ CML remission HSCs[253]. *GAS2* plays a role in various pathways including promoting cell proliferation and reducing apoptosis[254] which may be relevant for this JS+ enriched HSC_mpp2 cluster[254].

Although TNF via NFKB pathway gene signatures were overall downregulated in JS+ HSCs, this pathways was upregulated in HSC_mpp2 relative to HSC_mpp1. TNFA expression was increased in HSC_mpp2 (3.9%) as compared with HSC_mpp1 (1.2%). Within HSC_mpp2, more JS+ cells expressed TNFA (4.6% as compared with 1% in HD and 2% in J+, Figure 4.10G) and also in later JS+ progenitor populations (MkP and EBM) (Figure 4.10G). Hallmark interferon gamma response pathway and interferon-related genes were upregulated in HSC_mpp2. *IFI44L* was in expressed in 25.5% of HSC_mpp2 cells as compared with 13% of HSC_mpp1 cells, although this gene also remained upregulated in JS+ cells in HSC_mpp1 (Figure 4.10H).

Although cell-cycle and quiescence gene scores didn't differ significantly between the clusters, the expression of ADT CD133, a canonical HSC stemness marker[255], was reduced in HSC_mpp2 cluster which was largely specific to reduced expression in JS+ cells (Figure 4.10I). Megakaryocyte signature genes were increased in HSC_mpp2; VWF was expressed in 9.6% of HSC_mpp1 cells as compared with 28.4% of HSC_mpp2 cells (Figure 4.10J). Further, *TPSB2* was enriched in JS+ HSC_mpp2, expressed in 23.5% of cells, as compared with 13% of JS+ cells in HSC_mpp1, p<0.0001. Interestingly, MHC class II antigen genes were highly upregulated in HSC_mpp2 (Figure 4.10C, K). MHC class II antigens have recently been shown to be expressed in normal and leukaemic HSCs with antigen presenting capabilities[256]. The gene signatures in

JS+ enriched HSC_mpp2 may reflect multipotent progenitors rather than true HSCs considering reduction in stemness markers but also may reflect an MK- and potentially EBM-biased HSC cluster which is less quiescent and more proliferative.

Combinatorial ADT marker expression of CD34, HLA-DR/DP/DW, CD117, CD36, CD71 and CD235ab faithfully distinguished early to late-stage ErP (Figure 4.11A)[257]. JS+ ErP showed reduced heme metabolism signatures relative to J+ or HD ErP (Figure 4.11B). Specific to JS+ ErP, there was a reduction in expression throughout erythroid maturation but most evidently in later ErP in key haem synthesis genes including *ALAS2*, *AHSP* and *TSPO2* (Figure 4.11C). *KLF1,* a master regulator of erythropoiesis, was downregulated in JS+ ErP at all erythroid stages.

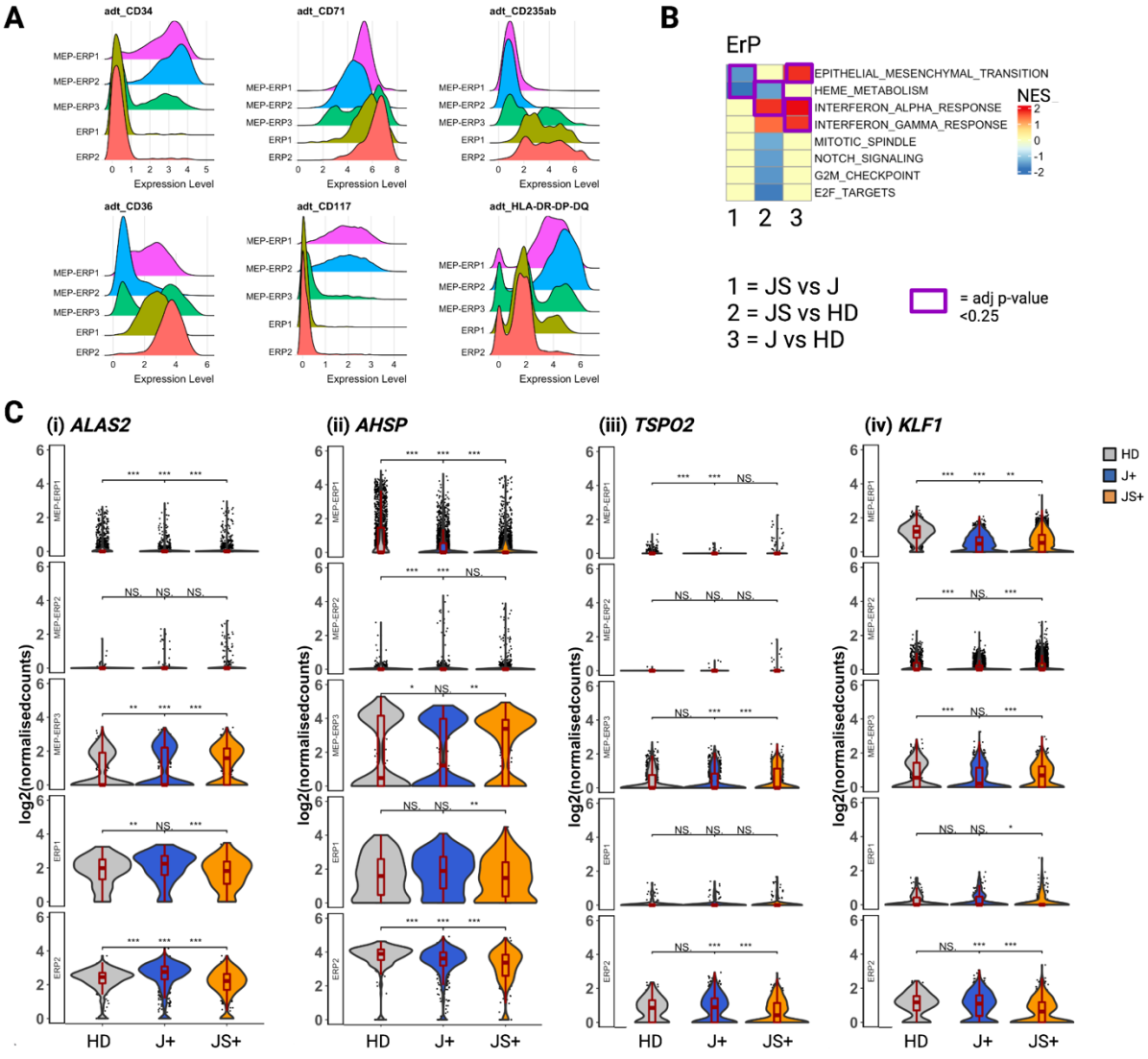


Figure 4.11. Differential gene expression analysis in 10X ErP. (A) Ridge plots showing canonical erythroid surface protein markers by ADT expression distinguishing ErP subpopulations with maturation

from early to late: MEP-ERP1→ MEP-ERP2→ MEP-ERP3→ERP1→ERP2. (B) GSEA analysis showed downregulation of haem synthesis pathways in JS+ ErP. (C) Violin plots of select genes illustrating reduction in specific ErP subpopulations: (i) *ALAS2*, (ii) *AHSP*, (iii) *TSPO2*, (iv) *KLF1*. Each dot represents the gene expression for each cell, boxplot in red represents median and quartiles. Wilcoxon test was used test statistical significance, comparison displayed left to right (HD vs. J+, HD vs. JS+ and J+ vs. JS+). NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

I analysed the protein expression of ADT markers in the HSPC populations to determine if there was a JS+ specific increase in expression of ADT markers particularly in the expanded progenitor populations (Figure 4.12A). There was downregulation in CD47 expression overall in MPN HSPCs as compared with HD HSPCs which was not distinguished by CD47 RNA expression (Figure 4.12B) and CD164 reduced in JS+ HSPCs overall. CD47 "don't eat me" protein is upregulated in normal HSCs and binds to SIRP-α expressed on macrophages to prevent HSC phagocytosis[258]. CD47 expression was reduced in a J+ MF cohort[259] but recently an analysis including PV, ET and MPN patients showed increased CD47 expression. However, this work was carried out in total MNCs which would mask HSPC-specific CD47 expression[260]. This reduction in expression may be relevant since a monoclonal anti-CD47 antibody[261], magrolimab, is showing efficacy in phase 2 and 3 trials for MDS and AML (including transformed MPN cases; NCT04435691). The efficacy of this drug could be postulated to be attenuated in J+ or JS+ patients due to lower CD47 expression.

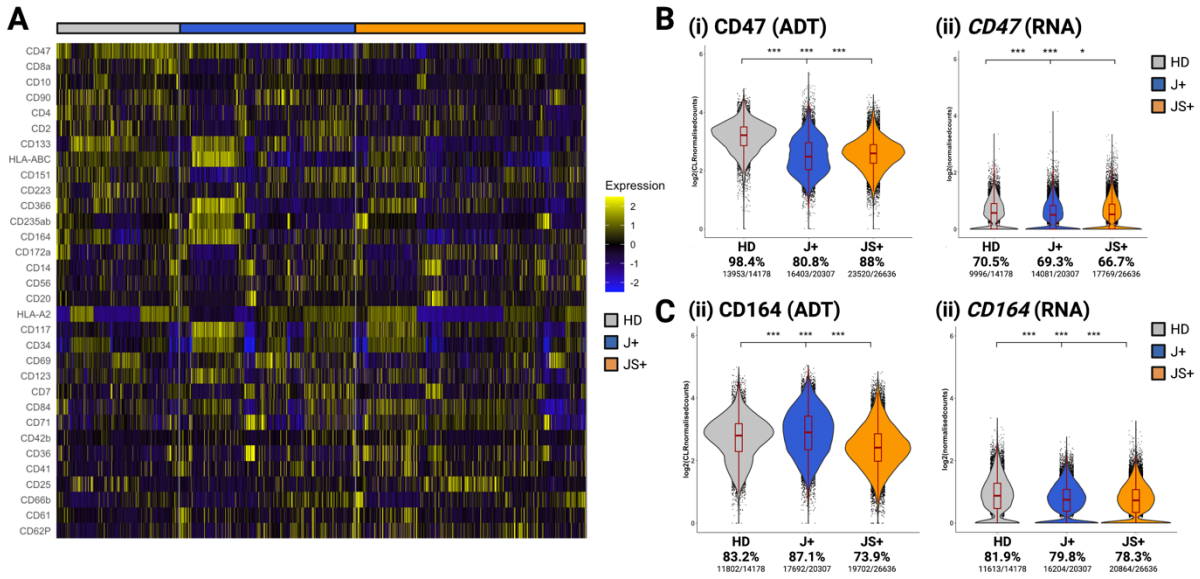


Figure 4.12. Differentially expressed ADT markers in 10X HSPCs. (A) Heatmap of differentially expressed ADT markers in HSPCs. (B) Violin plots showing CD47 expression by ADT (left) and RNA (right) by donor type (top panel) and CD164 expression by ADT and RNA by donor type (bottom panel). Each dot

represents the gene/ADT expression for each cell, boxplot in red represents median and quartiles. Percentage of cells expressing the gene presented below each plot. Wilcoxon test was used test statistical significance, comparison displayed left to right (HD vs. J+, HD vs. JS+ and J+ vs. JS+). NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

4-2.7 Innate and adaptive immune system in JS+ MPN

The bone marrow microenvironment comprises the cellular and structural components that are integral to the support and maintenance of hematopoietic cells. In MPNs, this microenvironment is disturbed[98] and a pro-inflammatory/fibrotic state which is permissive to promoting progression of the malignant clone[98,99]. The non-malignant HSPC cellular compartment constitutes both haematopoietic immune and non-haematopoietic cells (eg stromal cells and osteoblasts). The innate and adaptive immune systems are critical components which interplay to prevent cancer development and evolution, known as “immune surveillance” but importantly cancers can develop mechanisms to evade these systems[262].

In MPN, there is evidence this immune surveillance system is dysregulated contributing to disease phenotype and evolution. NK and certain T cell subsets have been reported to be reduced in MPN and MHC class I and II gene expression reduced in all MPN subtypes[263-265]. To explore the JS+ MPN mature immune compartment with finer granularity, I separated mature immune populations (macrophages, monocytes, T, B, NK and dendritic cells) from the overall object and repeated unsupervised Louvain clustering on 48,334 cells (Figure 4.13A).

After removing two contaminating HSPC clusters (Figure 4.13B) (n=2536), I performed integration comparing both RNA alone, and with multi-omic RNA and ADT integration (WNN) (Figure 4.13C, D). Delineation of specific cell types was improved with WNN integration. For example, NK subsets (early to mature) were resolved with WNN but not RNA integration alone (Figure 4.13, D). ADT markers of NK maturation identified the majority of NK clustered as CD56-CD16++CD57+ (mature NK), and two smaller clusters; an immature NK cluster CD56+CD16-CD57- and an intermediary maturing cluster CD56-CD16+CD57+(Figure 4.13E).

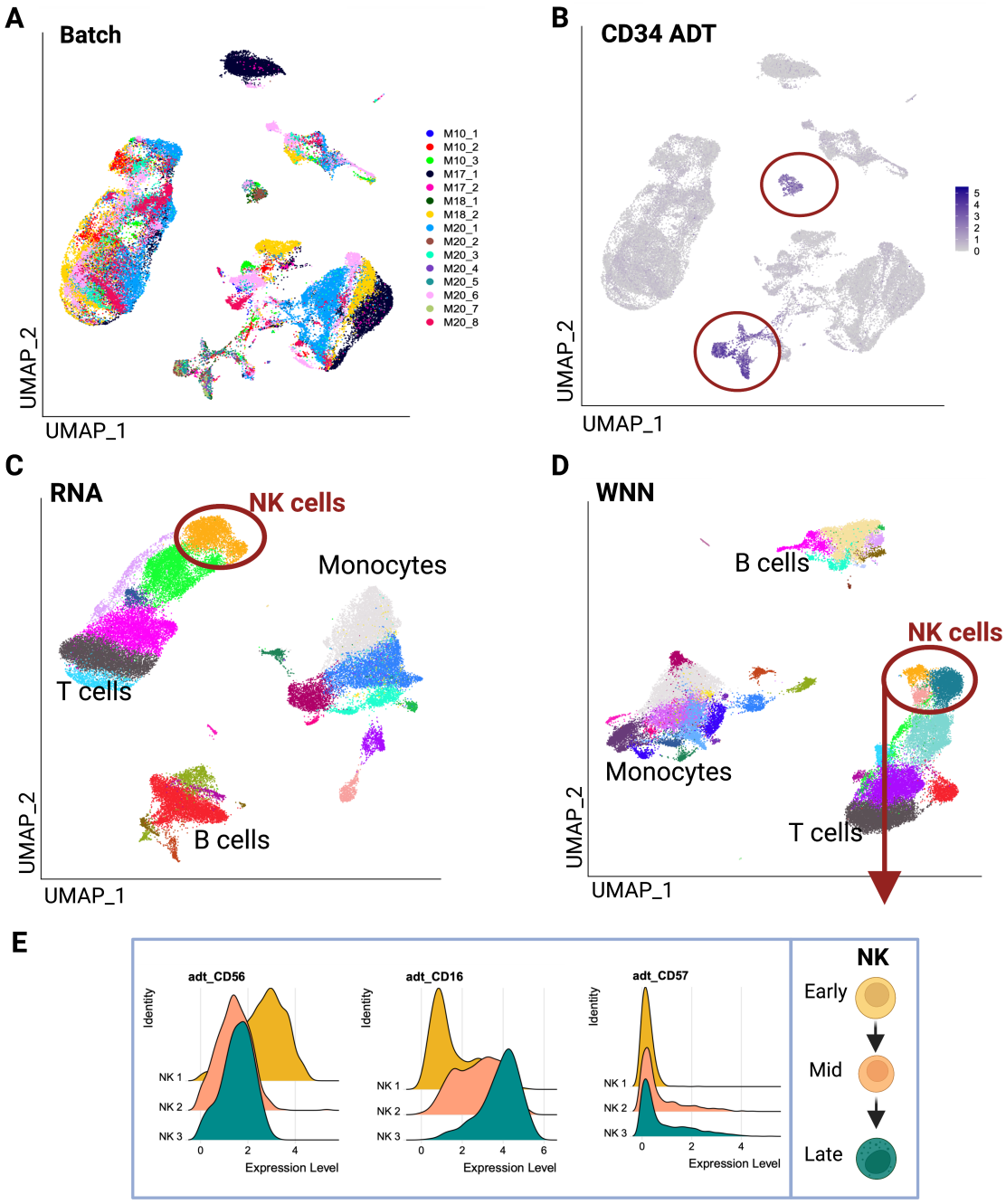


Figure 4.13. Mature cell population clustering. UMAP plots of mature cell populations coloured by batch (A) and highlighting CD34 ADT expression (B) showing HSPC clusters removed from downstream analyses. (C) UMAP plot of mature cell populations after HSPC removal, highlighting one NK cell population. (D) UMAP plot of mature cell populations clustered using WNN analysis highlighting 3 NK cell clusters which correspond to early, mid and late NK cells (E).

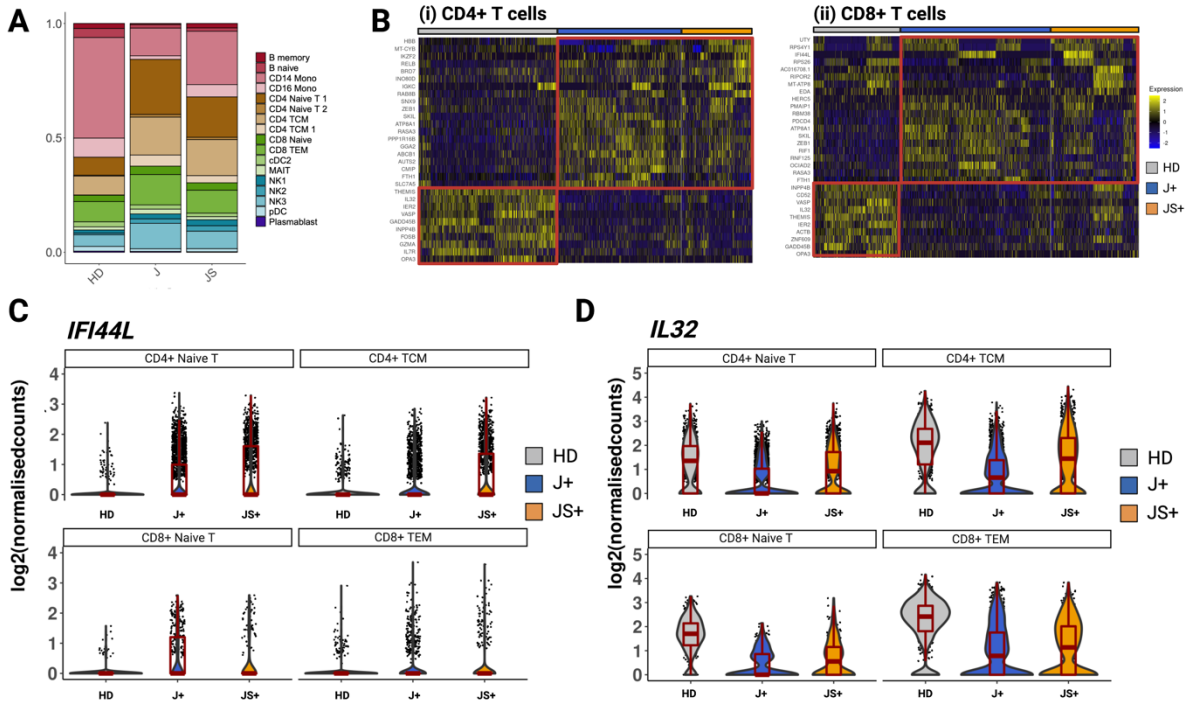


Figure 4.14. Differential gene expression analysis in immune cell compartments. (A) Barchart showing mature cell populations by donor type. (B) Heatmap of DGE in (i) CD4+ and (ii) CD8+ T cells between donor types. Violin plots showing (C) *IFI44L* and (D) *IL32* gene expression in CD4+ & CD8+ T cells by donor type. Each dot represents the gene expression for each cell and boxplot in red represents median and quartiles.

As shown in overall cell analysis, monocyte populations were depleted in MPN (JS+ and J+) as compared with HD (Figure 4.14A) with expansion of CD4+ T cell, CD8+ T cell and NK cell populations. I performed differential gene expression analysis between JS+, J+ and HD in CD4+ T cell, CD8+ T cell and monocytic/dendritic cell populations. MPN patient CD8+ and CD4+ T cells had distinct signatures as compared with HD CD4+ and CD8+ T cells (Figure 4.14B). Interferon signalling genes such as *IFI44L* were upregulated across MPN T cell subsets (Figure 4.14C) and there was downregulation of T cell activation markers such as *IL32* (Figure 4.14D).

MPN monocytes had a strikingly distinct gene signature as compared with HD monocytes (Figure 4.15A). These signatures were most prominent in CD14+ classical monocytes and still present but to a lesser extent in CD16+ non-classical monocytes. GSEA showed JS+ and J+ monocyte upregulation of multiple pro-inflammatory pathways; TNFA signalling, interferon alpha and gamma, TGFB signalling, inflammatory response, epithelial to mesenchymal transition and

IL2/STAT5 signalling pathways. (Figure 4.15B). MPN monocytes displayed downregulation of canonical monocyte markers including *VCAN*, *MNDA* (Figure 4.15C) and upregulation of interferon signalling genes (*IFNGR2* and *IFI27*; Figure 4.15D) and iron metabolism genes (*FTH1* and *HIF1A*; Figure 4.15E). Moreover, markers of fibrosis (*TIMP-1*[266], *TGFB1*[266], *CXCL8*[267]) and angiogenesis (*VEGFA*) were elevated in MPN classical monocytes (Figure 4.15F-G). Of the cells where *JAK2*V617F was detected, 6.9% were in monocytes and interestingly, only a minor contribution from other mature cell populations (combined frequency of 1.9%) with 91.2% detected in HSPCs.

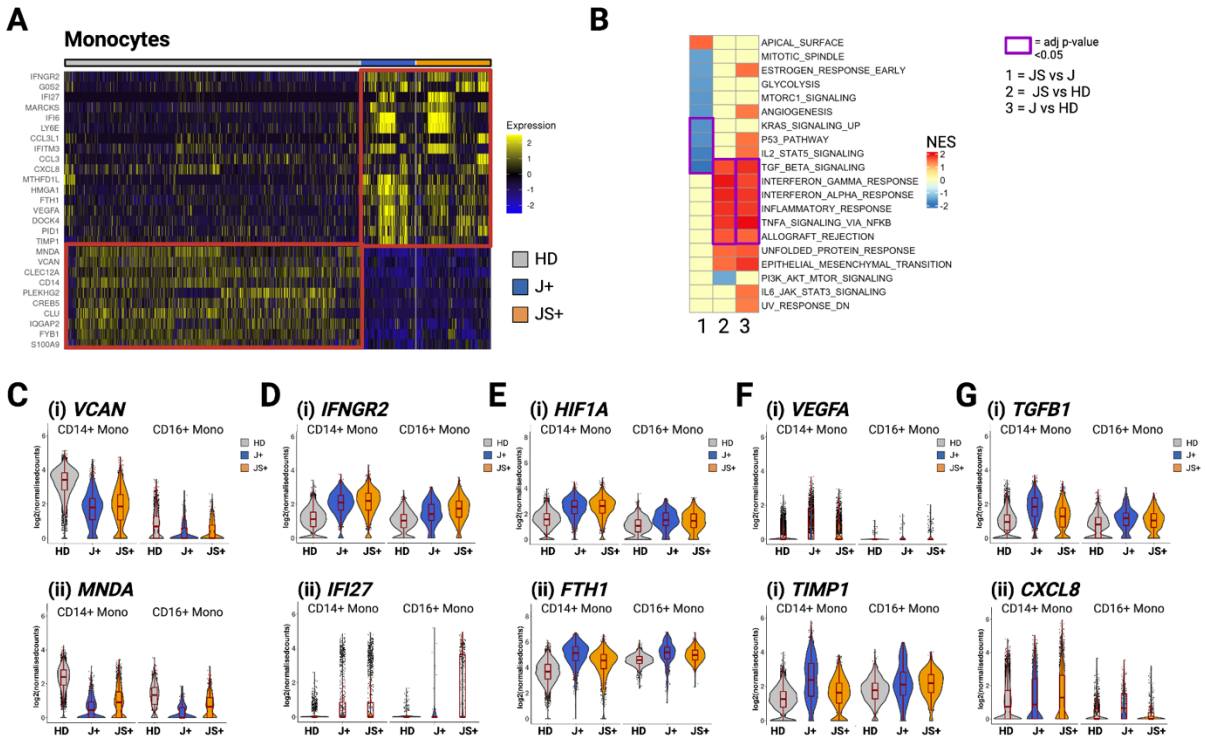


Figure 4.15. Differential gene expression analysis in monocytes. (A) Heatmap of DGE in monocytes between donor types. (B) GSEA analysis showing Hallmark pathways in monocyte populations with upregulation of multiple pathways in MPN monocytes. Violin plots showing expression of *VCAN* (C,i), *MNDA* (C,ii), *IFNGR2* (D,i), *IFI27* (D,ii), *HIF1A* (E,i), *FTH1* (E,ii), *VEGFA* (F,i), *TIMP1* (F,ii), *TGBF1* (G,i), *CXCL8* (G,ii) gene expression in monocyte populations by donor type. Each dot represents the gene expression for each cell and boxplot in red represents median and quartiles.

4-2.8 Regulatory gene network analysis

I performed regulatory network analysis using single-cell regulatory network inference and clustering (SCENIC) to determine the regulators driving JS+ aberrant expansion of HSPC cell populations and their associated molecular signatures, This method identifies regulons which are

genes that are known to be co-expressed with or targets of transcription factors. Regulon activity was expressed as an AUCell score. Initial interrogation of regulons revealed that cell types had distinct regulon profiles (Figure 4.16A).

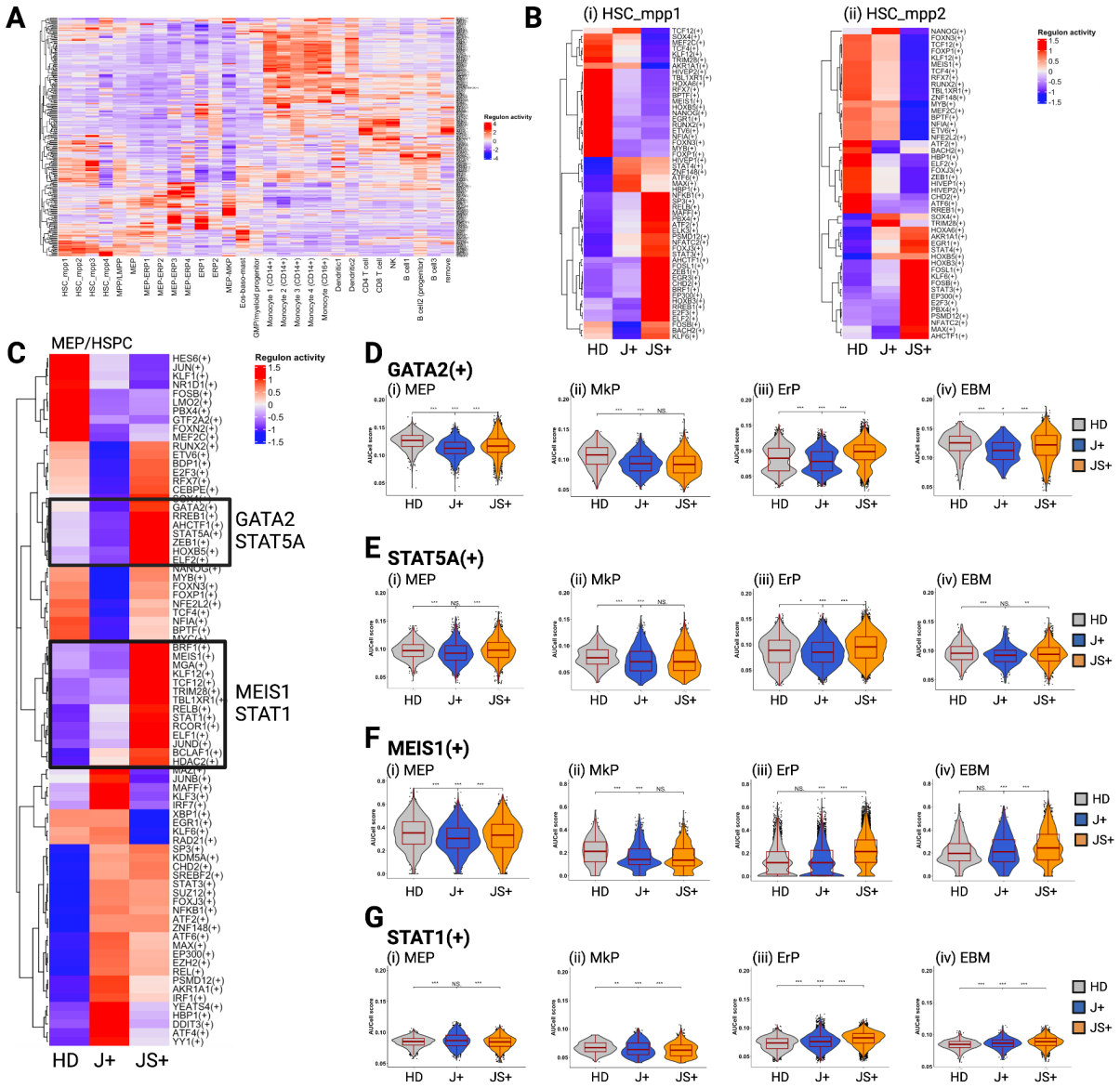


Figure 4.16. Regulatory networks in 10X HSPC populations. (A) Heatmap of regulon activity (expressed as AUCell score) by cell type showing that related cell types have similar regulon activity. (B) Heatmaps of regulon activity in (i) HSC_mpp1 and (ii) HSC_mpp2 clusters by donor type showing distinct regulon activity within each cluster. (C). Heatmap of regulon activity in MEP/HSPC clusters by donor type highlighting regulons upregulated in JS+ cells. Violin plots of select regulon AUCell scores which are upregulated in JS+ HSPCs. These are shown by donor and plotted for MEP, MkP, ErP and EBM populations illustrating *GATA2* (D), *STAT5A* (E), *MEIS1* (F) and *STAT1* (G) regulon activity. Regulons were identified with "(+)" after each transcription factor name.

Examining HSPC regulon activity by donor group showed distinct regulon profiles in each donor group in earlier HSCs and MEP/HSPCs (Figure 4.16B, C). There was upregulation of potentially relevant TF activity in JS+ MEP/HSPC populations. For example, lineage specification TFs including GATA2, MEIS1 and STAT TFs including STAT5A and STAT1 (Figure 4.16C) were increased in MEPs but these were principally enriched in JS+ ErP and to a lesser degree, EBM progenitors (Figure 4.16D-G). GATA2 is critical for mast cell and basophil development[268,269] and STAT5 is described to positively regulate GATA2 but is not essential for mast cell development[269]. MEIS1 enhances megakaryocyte-erythroid development and preferentially increases megakaryocytic rather than erythroid differentiation[270]. Upregulation of STAT TF activity is recognised as central to MPN pathogenesis including STAT1[45]. Therefore, increased STAT1 downstream targets in JS+ aberrantly expanded populations may be pertinent to the mechanism underlying the JS+ driven enhanced MPN phenotype.

4-3 Discussion

Multi-omic single-cell sequencing enabled characterisation of the haematopoietic landscape of JS+ MPN uncovering specific expansion of HSPC populations, namely MEP, ErP, MkP and uniquely EBM progenitors. These populations were expanded as compared with both *JAK2*V617F and HD populations, although *JAK2*V617F HSPCs were expanded relative to HD. The expansion in these rare EBM progenitors was an unexpected observation which could only have been achieved through studying single cells. In parallel, there was a reduction in mature cell populations but most strikingly in monocytes in both J+ and JS+ cells with distinct pro-inflammatory and pro-fibrotic molecular signatures.

Single-cell analyses of *CALR*+ and *JAK2*V617F MPN HSPCs have already demonstrated aberrant haematopoiesis with megakaryocyte progenitor expansion[124,271] and associated dysregulated gene signatures, for example unfolded protein responses specific to *CALR*mut ET and MF MkPs[271]. Proliferation and fibrosis signatures were associated with *CALR*+ and *JAK2*V617F MF MkPs[124] which is in line with earlier data proffering MkPs as central drivers of fibrosis in MPNs[229,230]. My analysis replicates MkP progenitor population changes in MPN.

Although enhanced erythropoiesis is pathognomonic of PV[272], single-cell analyses in MPN haven't explicitly described abnormal erythroid progenitor expansions. Increased but dysfunctional

erythropoiesis is a hallmark feature of *SF3B1*mut MDS[273,274] and in particular, abnormal terminal erythroid differentiation[275,276]. A recent preprint *SF3B1*mut MDS single-cell analysis observed a mutant-specific expansion of MEP and erythroid progenitors[127]. In my analysis, these populations were most notably increased in JS+ samples but were also to a lesser extent in J+ samples. Haem synthesis pathways were downregulated in JS+ ErPs which may reflect abnormal effective terminal differentiation.

EBM progenitors were increased 3-fold in JS+ as compared with J+ and HD which hasn't been reported in *SF3B1*mut MPN. Mast cells have been reported to be increased in small studies of MPN[277-279] but not in recent MPN single-cell analyses. An image analysis of tryptase-stained bone marrow biopsies from MPN patients (n=31) found ET (n=4) and PMF (n=6) cases had significantly more mast cells than CML, PV or controls and mast cell proportion correlated with bone marrow fibrosis[277]. These studies didn't correlate mast cell frequencies with MPN genomic information. Murine models of *JAK2*V617F and *MPL*W515L MPN recently reported an increase in mast cells progenitors in single-cell RNA sequencing and IHC[280]. Mast cells have been associated with fibrosis development in benign (pulmonary[234], cardiac[235], renal[236] and liver[237]) and malignant conditions[238] and are considered central players in driving fibrosis[239]. Tryptases were among the most highly differentially expressed genes in JS+ HSPCs in my dataset including the earliest HSCs suggesting potentially "EBM-primed" HSCs in JS+ MPN. Tryptases have been directed implicated in promoting fibrosis[281] and notably in the KIT-mutated myeloid neoplasm, systemic mastocytosis, patients often develop fibrosis[282]. As well as upregulation in tryptase expression, other genes which may be pertinent in fibrosis development were dysregulated specifically in JS+ HSPCs such as *CLEC11A* which encodes for osteolectin or stem cell growth factor.

In early HSC/MPP populations, two distinct HSC clusters were present, one of which was dominated by JS+ cells. This cluster displayed a distinct molecular program; upregulation in MK and EBM lineage signature genes, IFN signalling genes and downregulation in apoptosis genes and canonical stemness markers (eg. CD133). Analyses of *JAK2*V617F HSPCs from patients with ET and PV observed an expansion of MK-biased HSCs[283,284] and upregulation in IFN response genes in these populations[283]. These signatures suggest this may be a MK/EBM-primed HSC cluster driven primarily by JS+ HSCs. I hypothesise that these HSCs drive the expansion of the abnormal HSPC progenitors in my dataset.

Understanding how *SF3B1*-mutant aberrant splicing drives the cellular and molecular changes I have described is essential. However, studying this was not possible in the 10X dataset for a number of reasons. Firstly, determining mutant-specific splicing events is not feasible with this approach due to low sensitivity of mutation detection. In my dataset, *JAK2*V617F and *SF3B1* mutations were only detected in 1-2% of the reads. Secondly, this method is 3' biased, short read and the libraries were sequenced to a depth insufficient to adequately perform differential splicing analysis. Splicing event detection plateaus at 1-2 million reads per cell and cells were sequenced to ~50,000 reads per cell. The cost to sequence each cell at the depth required for meaningful splicing analysis would be prohibitive and nevertheless, full length transcriptome information is not available with this method. These limitations can be overcome with TARGET-seq which can detect mutant and non-mutant cells with high sensitivity in parallel with full-length transcriptome permitting study of mutant-specific splicing signatures. I applied this method to the dataset and explore these results in chapter 5.

Neoantigen detection through 10X dataset is more challenging in this dataset for the discussed limitations in detecting mutant-specific splicing events and TARGET-seq has the potential to circumvent this. Dysregulation in the MPN immune compartment, eg an exhausted T cell phenotype, may prevent an adequate response and clearance of cells expression neoantigens. The T cell compartment in my dataset, however, didn't display molecular signatures of an exhausted immune phenotype.

Overall, this single-cell multi-omic analysis demonstrates that *SF3B1* mutation acquisition promotes an enhancement of *JAK2*V617F MPN phenotype with expansion of specific HSPC subpopulations (MkP, ErP and EBM) and molecular signatures in earlier HSC/MPP suggestive of enhanced lineage priming of these populations.

Chapter 5: Mutant-specific molecular signatures and aberrant splicing events in HSPCs of *JAK2*V617F-*SF3B1*-mutated MPN

5-1 Introduction

The canonical *JAK2*V617F mutation has been detected in HSPCs from the apex of the hierarchy, the HSC and in the context of MDS, *SF3B1* mutations have also been identified in HSCs[228]. In JS+ MPN, the clonal hierarchy has yet to be precisely elucidated. The order of mutation acquisition, double mutant as compared with single mutant HSPCs and interrogation of WT HSPCs may be significant for underlying disease biology. Order of mutation acquisition can influence MPN disease phenotype. When *JAK2*V617F and *TET2* mutations are present, single-cell colony experiments have illustrated that a "JAK2-first" scenario was more often associated with a PV phenotype whereas "TET2-first" was associated with an ET phenotype[94]. Similar observations occur for *DNMT3A* and *JAK2*V617F mutations where again "JAK2-first" and "DNMT3A-first" patients had a PV and ET phenotype, respectively[95].

Plate-based single-cell genomics approaches such as TARGET-seq are ideal for resolving genetic intratumoral heterogeneity including order of mutation and clonal hierarchies of mutations in MPN; for example demonstrating that spliceosome mutations typically occur before MPN driver mutations[122] and importantly also in distinguishing WT HSPCs. This approach also can establish zygosity of mutations within individual cells, which is relevant for *JAK2*V617F in MPN. This method was recently applied to a *TP53*-mutant MPN cohort determining the central role WT HSCs play in disruption of haematopoiesis with increased inflammatory pathway gene signatures detected in WT HSPCs[122]. Mutant-specific gene signatures can be identified; previous analysis elucidated cell surface antigens such as *G6B* specifically upregulated in mutant HSPCs in MF[284], underscoring the strength of single-cell technologies to discover the signatures associated with disease-driving populations and uncover novel treatment targets[124].

Importantly for studying JS+ MPN, TARGET-seq can be used to identify mutant-specific splicing events through high-sensitivity genotyping coupled with full-length transcriptome sequencing of each single cell. This is essential to understand mutant-specific splicing changes that may be implicated in the disease biology.

Aberrant splicing events in *SF3B1*mut myeloid neoplasms have been predominantly studied at the bulk level[285,286] (Figure 5.1A-E) with introduction of a cryptic A3SS the most prevalent event described[222,273,285,287-291] (Figure 5.1C, 5.2). A3SS of *MAP3K7* has been one of the most extensively studied event in MDS[273]; most recently, downregulation in *MAP3K7* transcription with reduced protein expression was shown to result in premature *GATA1* downregulation, an essential regulator of erythropoiesis, with accelerated erythropoiesis and erythroid cell apoptosis[273]. Furthermore, aberrant splicing of mitochondrial transporters, *ABCB7* and *TMEM14C*, drives iron sequestration causing ring sideroblast development in dysfunctional erythroid cells in *SF3B1*mut MDS[292]. Another *SF3B1*mut ASE relevant to understanding the mechanisms driving *SF3B1*^{mut} cancer development *is an* A3SS-induced aberrant isoform of *PPP2R5A*[289]. In the normal context, this gene codes for a post-translational regulator of MYC and in *SF3B1*^{mut} malignancies, an abnormal *PPP2R5A* isoform is susceptible to nonsense-mediated decay and reduced protein expression which leads to increased *MYC* phosphorylation and *activation.*

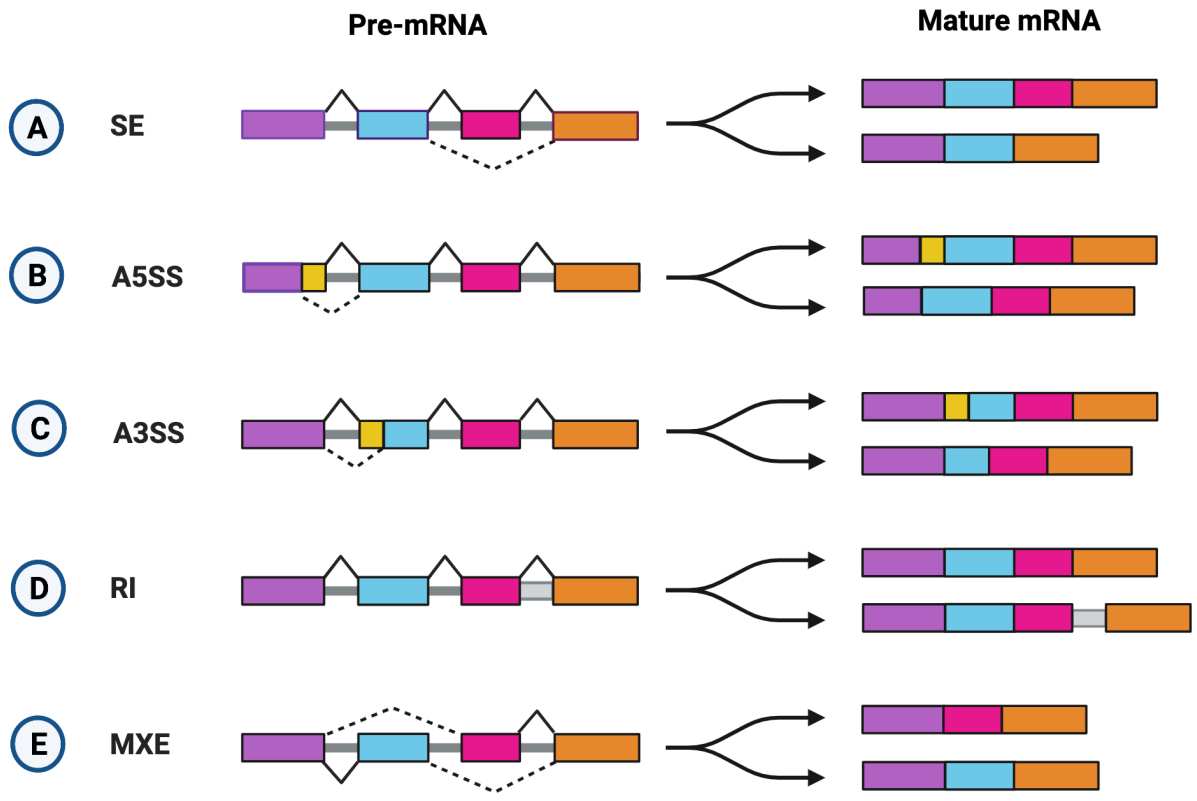


Figure 5.1. Alternative splicing events. (A) Skipped exon event (SE). (B) Alternative 5' splice site event (A5SS). (C) Alternative 3' splice site event (A3SS). (D) Retain intron event. (E) Mutually exclusive exon event.

Other ASE have been described in association with *SF3B1* mutations, including retained intron[226,285,293], exon inclusion[294] and less often A5SS[293] and skipped exon events[293]. Increase in excluded RI events have been described in splicing factor-mutated myeloid malignancies (including *SF3B1*mut)[293] and also *SF3B1*mut CLL[226]. Another consequence of *SF3B1*mut observed has been inclusion of a 'poison' exon through recognition of a deep intron branchpoint sequence, affecting *BRD9* gene[294]. this poison exon disrupts the reading frame with the transcript vulnerable to NMD and subsequently reduced *BRD9* expression levels.

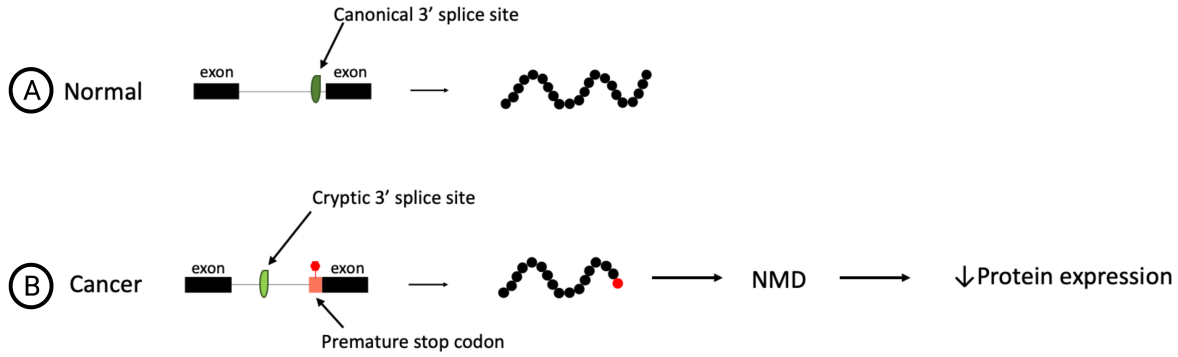


Figure 5.2. Schema of *SF3B1*mut-associated A3SS. (A) Canonical 3' splice site in normal circumstances. (B) *SF3B1*mut introduction of upstream cryptic 3' splice site which leads to inclusion of a premature stop codon and a shortened transcript which is subject to nonsense mediated decay (NMD) in ~50% of transcripts with reduced protein expression. Figure created in collaboration with Wei Wen.

The majority of *SF3B1*mut ASE have been identified using bulk RNA sequencing approaches and although the field of single-cell splicing analysis is relatively nascent, methodologies in both wet lab and computational methods to study this are being established at a rapid pace. GoT-splice[127] is a single-cell multi-omic approach (preprint published on bioRxiv, a preprint repository) studying aberrant splicing in *SF3B1*mut MDS at the single-cell level. The methodology applied is an advancement on currently available droplet-based multiomic approaches through adaptation of Genotyping of Transcriptomes (GoT)[271], a method where 10X genomics scRNA-seq is modified to include targeted primers at the first cDNA step to integrate genotyping information with transcriptome. GoT-Splice further expands on this by successfully detecting full-length transcripts through performing Nanopore sequencing on 10X libraries and finally combining these with CITE-

seq to obtain parallel genotyping, cell surface proteomic, full-length transcriptomic and splicing information for each cell. Nanopore technology[30] sequences a single molecule of DNA by passing it through a nanopore membrane under electric charge capturing the entire length of the transcript and sequencing is performed without PCR amplification.

GoT-splice was showcased in *SF3B1*mut MDS finding that *SF3B1* mutations were present throughout the haematopoietic hierarchy and that the mutations were increased in the erythroid trajectory. Splicing analysis results were consistent with that previously reported, cryptic 3' splice sites accounting for 57% of events and many events previously reported were identified, although not previously in a cell-specific manner. However, only 50% of the cells were successfully genotyped and the allelic dropout rate was not reported which is an important confounder in scRNAseq[295]. *SF3B1* mutations can only occur in heterozygous state therefore mutated cell calls were true positives whereas for WT cell calls, due to the sensitivity reported in this study, it is possible mutant cells were miscalled as WT which may confound accurate interpretation of the results.

TARGET-seq can overcome this limitation by combining gDNA and mRNA genotype information to assign zygosity with high confidence and an allelic dropout rate of ~5%. In *SF3B1*mut MPN, a concomitant JAK-STAT signalling mutation is detected in most cases, which can exist in a heterozygous or homozygous state. Therefore, a method that can accurately determine allelic state is a prerequisite. The high sensitivity genotyping that can be achieved with TARGET-seq can reliably distinguish mutant and WT cells which uniquely permits intra-patient comparison of these cells and identification of mutant-specific gene and splicing signatures.

5-1.1 Hypothesis

Single-cell methodologies distinguishing mutant and WT cells have the potential to uncover mutant and cell-specific aberrant gene and splicing signatures. MPNs typically develop following acquisition of a somatic mutation in a single HSC[6,7,25] which develops a clonal advantage and causing expansion of downstream progenitors (eg. MkP and ErP). Therefore, in JS+ MPN, I hypothesised that by understanding the mutant disease-driving early stem and progenitor cells through applying these methodologies, I would gain insights into the mechanisms promoting disease progression and ultimately identify tractable mutant and cell-specific therapeutic targets.

5-1.2 Aims

I aim to achieve this by applying TARGET-seq to:

1. Delineate JS+ MPN HSPC clonal hierarchy
2. Uncover JS+ specific molecular signatures in MPN
3. Understand whether there are aberrant splicing events affecting the function of known dysregulated pathways in MPN (eg. TNFA or NF-κB signalling)?

5-2 Results

5-2.1 Clonal hierarchy and heterogeneity of JS+ MPN

The cohort for this analysis consisted of 5 patients per mutation group (JS+ n=5, J+ n=5) and 5 healthy donors (HD), using one HD as a "batch control". The same healthy donor batch control was included in every plate which was required to control for technical variation such as the effect of different plates and sample processing times. Disease phenotype ranged from earlier chronic phase MPN (ET) to a more advanced MPN stage (PMF) (Table 4.1, Chapter 2). All JS+ cases were selected for the same *SF3B1* K666 hotspot mutation which is found more frequently in MPN as outlined in chapter 3. Single HSPCs were sorted into 384-well plates. I specifically sorted CD34+CD38- and CD34+CD38+ cells. Cells were index sorted meaning full immunophenotypic information (see Table 2.2, chapter 2) was recorded for each cell. 5716 cells were sequenced with 85% (n=4862) of cells passing QC.

The HSPC clonal architecture of each JS+ patient was heterogenous and order of mutation would have been inferred incorrectly by interpretation of PB bulk mutation level VAF alone (Figure 5.3). Where multiple mutations co-exist, the bulk mutation VAF may be a surrogate for mutation order acquisition where the mutation with a higher VAF is inferred to be the founding mutation. In 4 of 5 JS+ MPN patients, the most frequent HSPC genotype was co-mutation of *JAK2*V617F and *SF3B1* in the same cell. *SF3B1* was always present in the heterozygous state as expected. In

double mutant cells, *JAK2*V617F was present in both heterozygous and homozygous states. Unexpectedly, in one patient TNO123, *JAK2*V617F and *SF3B1* mutations resided in separate clones with 56.6% of the cells heterozygous for *SF3B1*, 29.5% of the cells were WT and a smaller proportion, 9.5% and 4.3% of the cells were heterozygous and homozygous for *JAK2*V617F respectively. This is a good example of a clonal hierarchy revealed by single-cell analysis that could not be resolved with bulk genotyping. This patient had ET with a bone marrow fibrosis WHO score 0 in contrast to the other 4 cases who had WHO fibrosis scores ranging 1 – 3. WT cells were identified in 3 of 5 JS+ patients at a frequency ranging 3 – 29.5%.

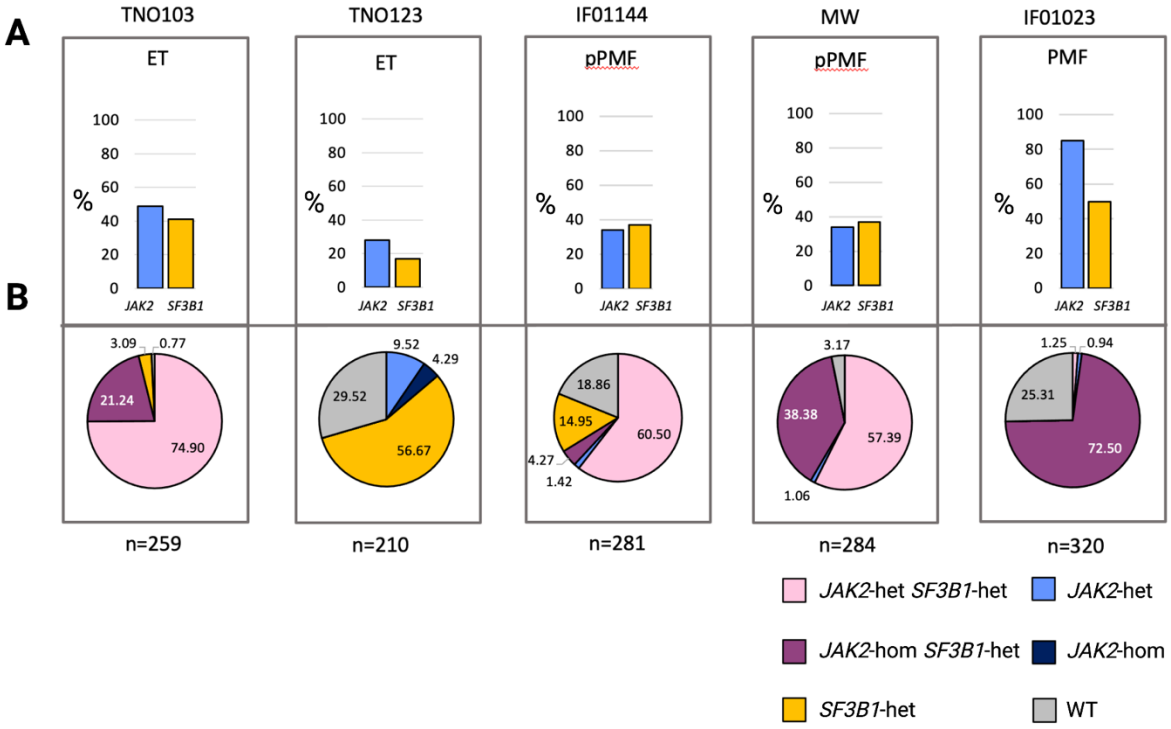


Figure 5.3. Clonal composition of JS+ MPN samples. (A) Peripheral blood bulk mutation VAF for each patient. (B) Pie charts of HSPC single-cell genotype showing proportion of each genotype detected at the single-cell level per patient (frequency in percentage shown on pie chart).

In J+ donors, as expected[296], heterozygous and homozygous *JAK2*V617F genotypes were detected in all cases. with the highest proportion of mutant HSPCs present in the patient with the most advanced MPN phenotype (IF01132; post-polycythaemia myelofibrosis), Figure 5.4. The PB VAF of *JAK2*V617F only donors could not be used as a surrogate of the HSPC compartment clonal burden. For example, patients TNO168 and IF01132 had similar PB VAF ~40%. However, in the HSPC compartment, 46% (TNO168) and 87% (IF01132) of the cells carried a *JAK2*V617F mutation. One of the healthy controls had *JAK2*V617F detected in 6% of cells. Explanations for

this result was either false positive due to contamination or true positive due to clonal haematopoiesis. I determined that this was a true positive result through orthogonal confirmation of *JAK2*V617F mutation detection in the 10X dataset for this control and no other *JAK2*V617F cells were detected for the other healthy controls in either dataset. Furthermore, blank wells were included in the experimental design showing no evidence of contamination. WT HSPCs were more prevalent in J+ patients ranging 13-79%.

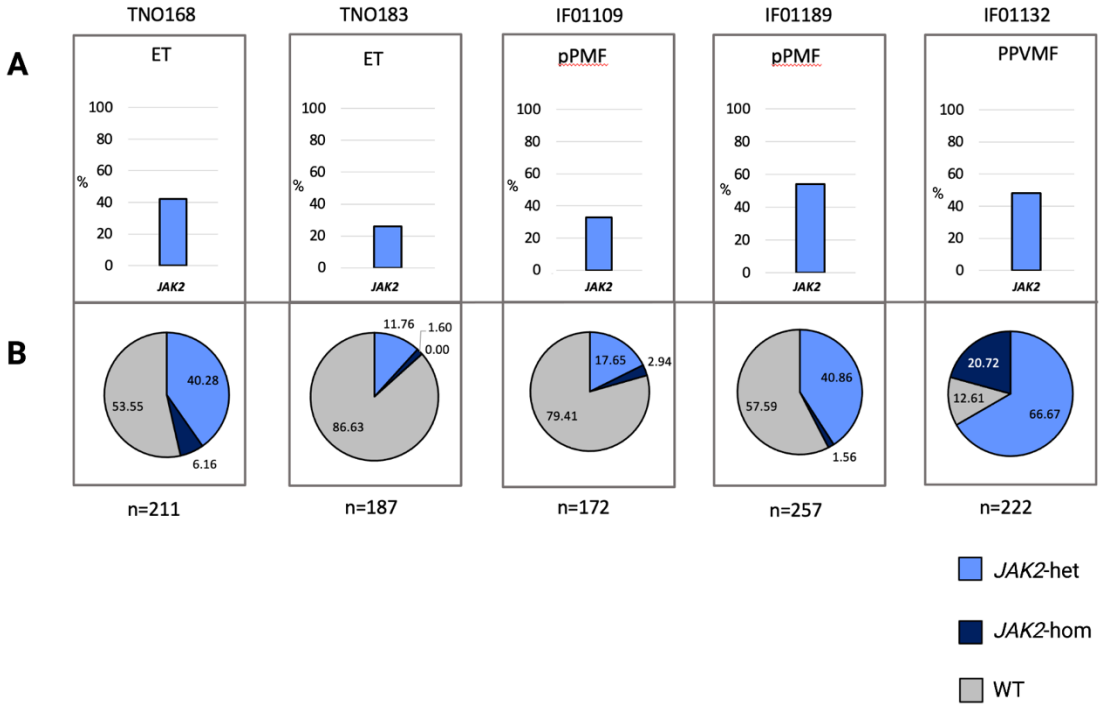


Figure 5.4. Clonal composition of J+ MPN samples. (A) Peripheral blood bulk mutation VAF for each patient. (B) Pie charts of HSPC single-cell genotype showing frequency of each genotype detected at the single-cell level per patient (frequency in percentage shown on pie chart).

The order of mutation acquisition in JS+ cases could be confidently established in two cases with *SF3B1* acquired first with evolution to *JAK2*V617F heterozygous and homozygous double mutant cells (Figure 5.3, TNO103 and IF01144). This reproduces previous findings that splicing factor mutations are more often initiating events[122,297]. For patient TNO103, the PB VAF of *JAK2*V617F was higher than *SF3B1* highlighting that PB VAF cannot be reliably used to infer mutation order. Determining the initiating mutation was not possible for two patients (Figure 5.3, MW and IF01023) where the majority of HSPCs carried both mutations. HSPCs sequenced were a mixture of HSCs, CD34+CD38-, CD34+CD38- and lineage negative CD34+ progenitors. I analysed the

index sorting information for each cell and In JS+ patients, both mutant and WT HSPCs were present throughout the HSPC compartment from the earliest HSC.

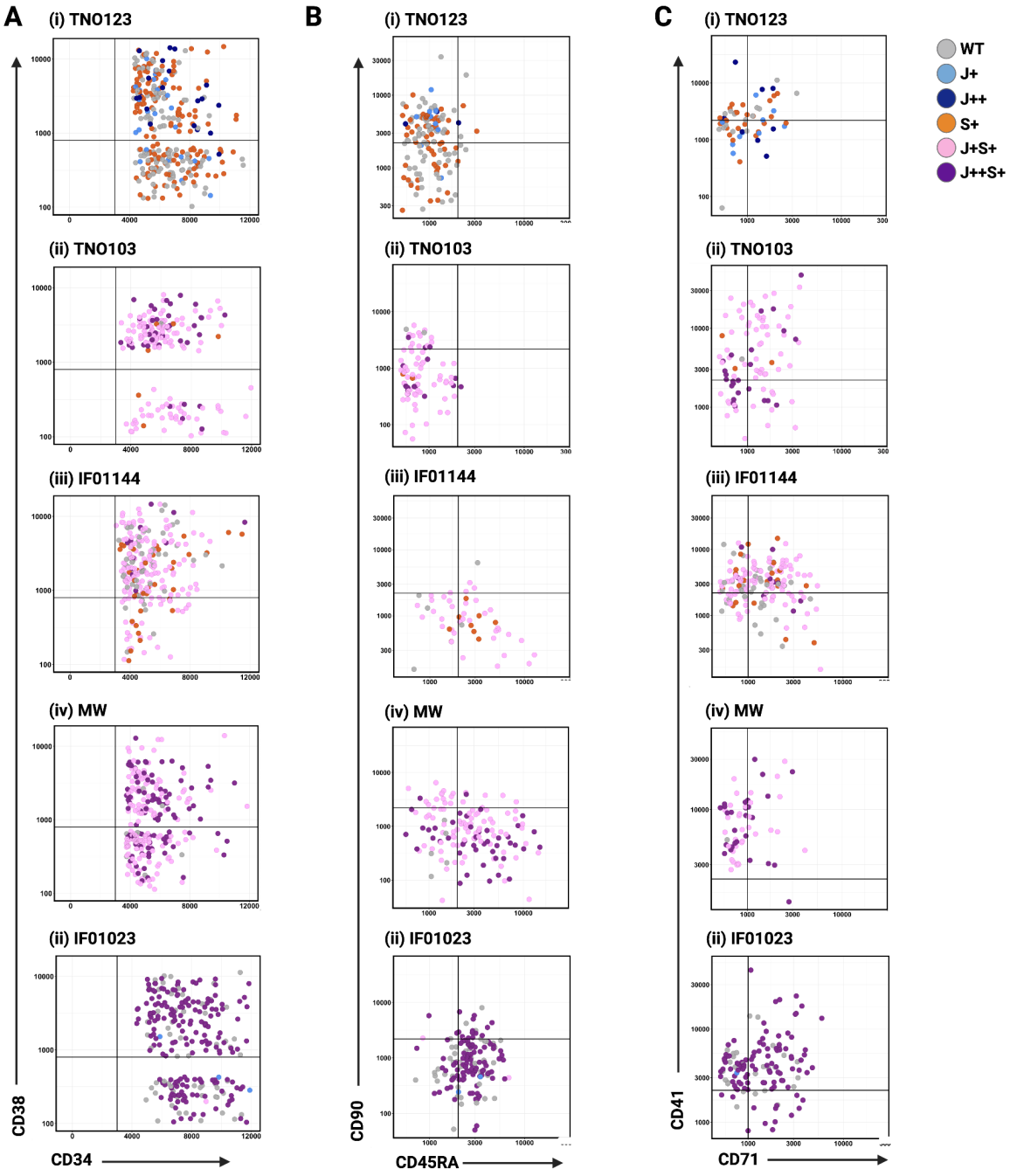


Figure 5.5. Integrated TARGET-seq genotyping and HSPC immunophenotype of JS+ HSPCs. Scatter plots of CD34+ vs. CD38+ expression (A). Scatter plots gating on CD38- cells only showing (B) CD45RA+ vs. CD90+ expression (B) and gating on CD38+ cells only showing CD71+ vs. CD41+ expression (D). WT=WT, J+= *JAK2*V617F heterozygous, J++= *JAK2*V617F homozygous, S+=*SF3B1* heterozygous, JS+=*JAK2*V617F heterozygous *SF3B1* heterozygous, J++= *JAK2*V617F homozygous.

There was no enrichment of a mutation in a specific progenitor cell type (defined immunophenotypically) for the majority except for TNO123 where HSPCs carrying homozygous *JAK2*V617F mutations were more prevalent in CD34+CD38+ compartment, Figure 5.5A (i). In J+ HSPCs, both heterozygous and homozygous clones were detected throughout the HSPC compartment (Figure 5.6).

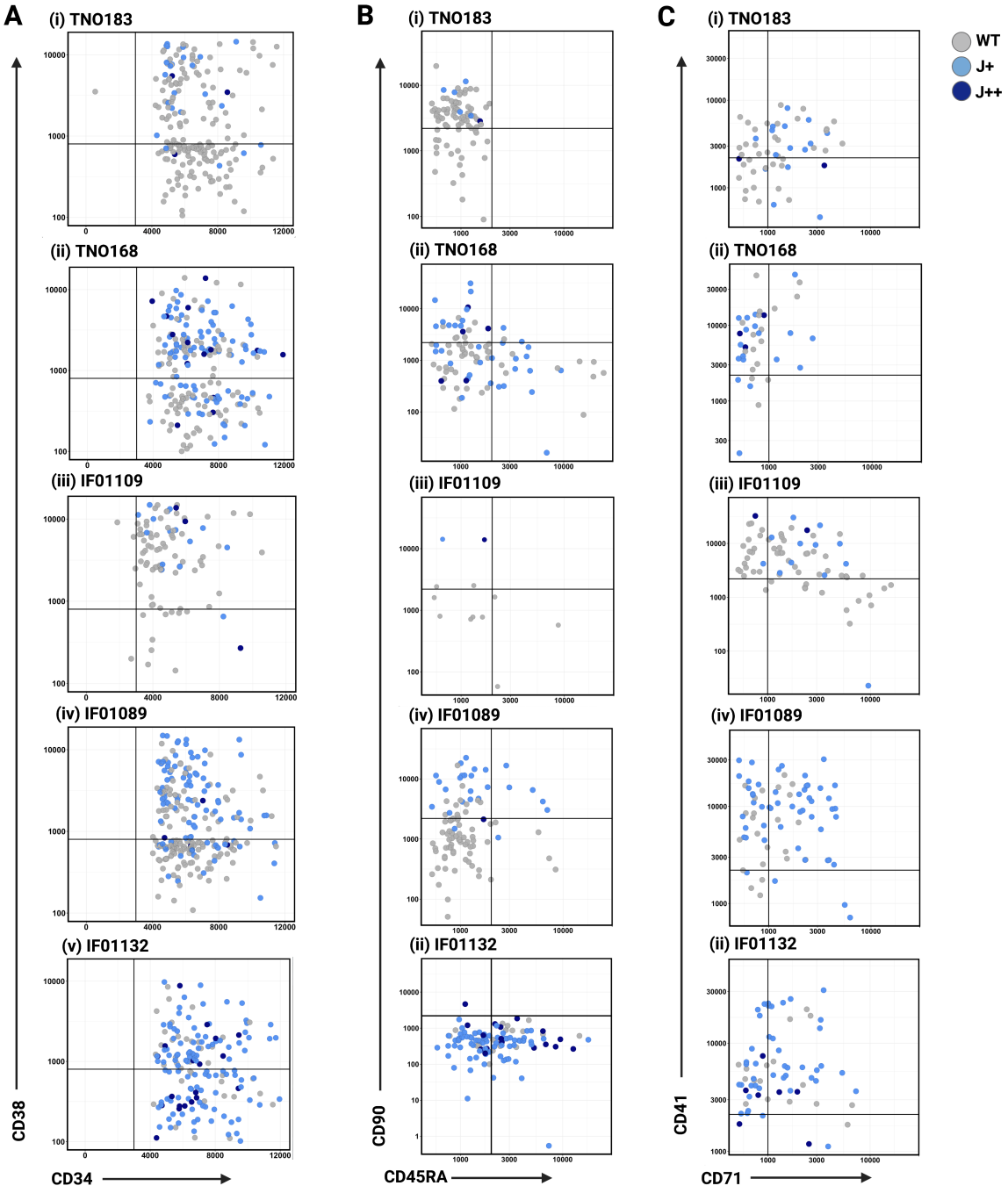


Figure 5.6. Integrated TARGET-seq genotyping and HSPC immunophenotype of J+ HSPCs. Scatter plots of CD34+ vs. CD38+ expression (A). Scatter plots gating on CD38- cells only showing (B) CD45RA+ vs. CD90+ expression (B) and gating on CD38+ cells only showing CD71+ vs. CD41+ expression (D). WT=WT, J+= *JAK2*V617F heterozygous, J++= *JAK2*V617F homozygous.

5-2.2 Transcriptomic profiles of JS+ MPN

Based on genotype profiles of the cohort, 2304 cells were selected for full-length transcriptomics. In total, 1,617 (70%) cells passed QC for analysis (Figure 5.7A-E). I selected 3599 highly variable genes for downstream gene expression analysis.

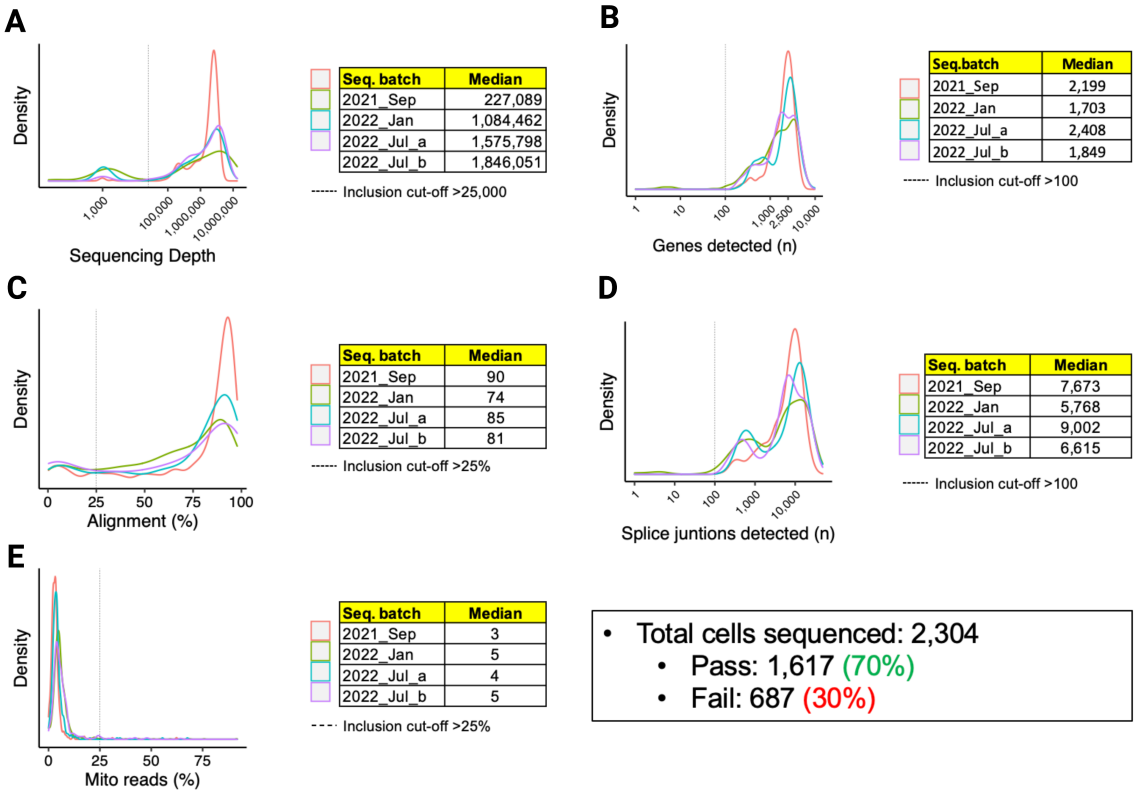


Figure 5.7. TARGET-seq quality control metrics. Density plots across all batches showing median sequencing depth (A), median genes detected per cell (B), median alignment percentage per cell (C), median number of splice junctions detected per cell (D) and median mitochondrial reads per cell (E). Pre-processing steps performed by Wei Wen, DPhil student in the Mead and Thongjuea laboratories.

Merging the cells across sequencing runs without integration showed some batch effects from sequencing batch (Figure 5.8A) and to a lesser extent sort run (Figure 5.8B) with consequent less optimal clustering of donors (Figure 5.8C) and genotypes (Figure 5.8D). Batches were integrated using Harmony with sequencing batch included as a covariate for regression (Figure 5.9A). Whilst there was some separation of cells carrying both mutations from WT and single-mutant cells after

harmony integration, cells didn't cluster discretely by genotype through unsupervised dimensionality reduction and clustering (Figure 5.9B). This replicates previously single-cell HSPC MPN transcriptome data[122] and illustrates the unique advantage of TARGET-seq to confidently distinguish mutant cells and their associated signatures from non-mutant cells, which would otherwise not be possible through unbiased approaches.

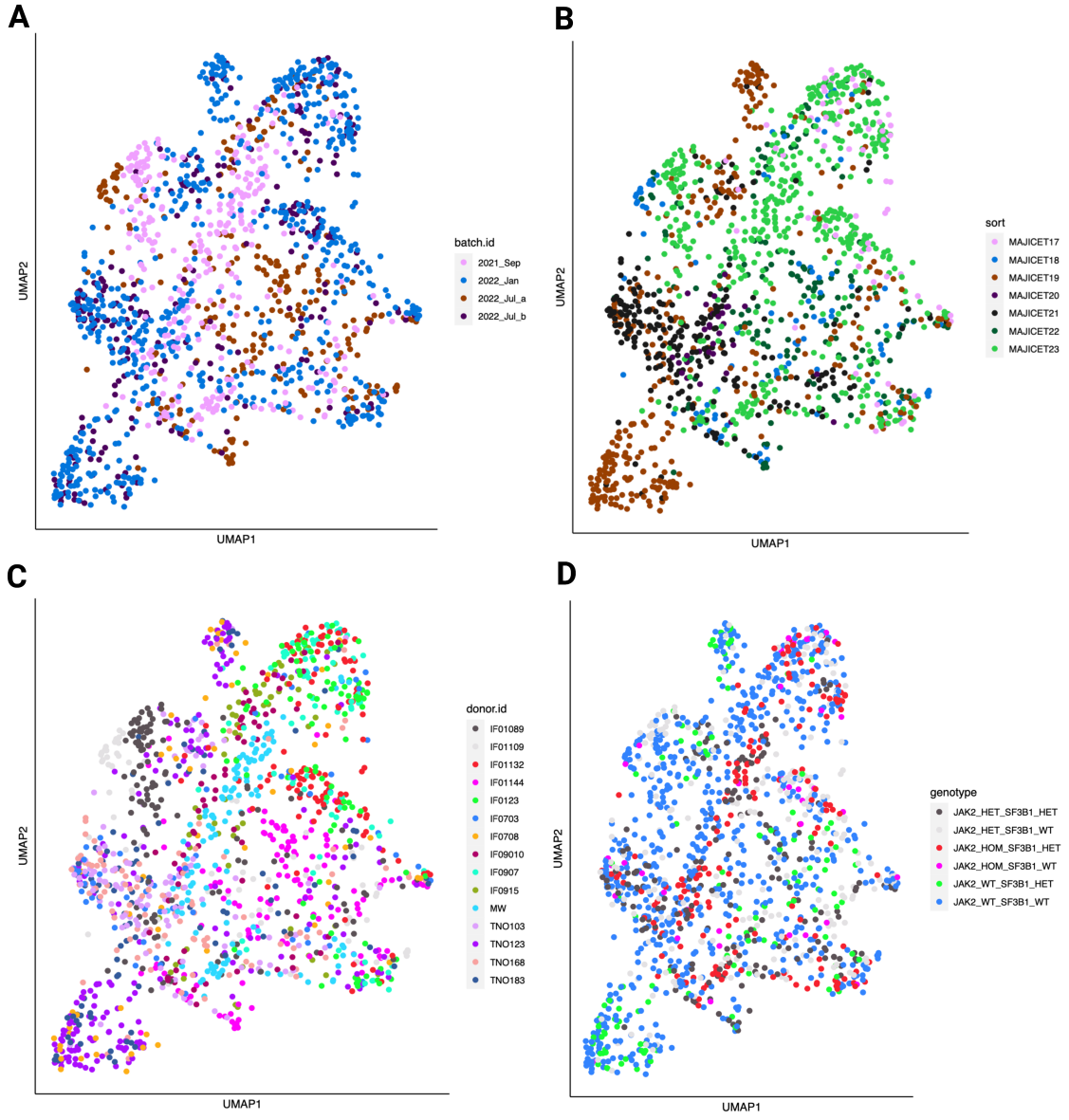


Figure 5.8. UMAP plots of TARGET-seq HSPCs merged without integration. UMAP plots coloured by sequencing batch (A), sort run (B), individual donor (C) and single-cell genotype (D).

Using FDG, I determined lineage differentiation trajectories to study the relationship between immature and more mature HSPCs and the presence of aberrant or expanded lineage

differentiation. I performed louvain clustering finding 11 HSPC clusters and ordered the cells on a force-directed graph. The majority of clusters were HSC/MPP and three clusters showed distinct lineage differentiation (Figure 5.9D). Clusters 8, 9 and 11 corresponded to erythroid, myeloid and megakaryocyte clusters respectively (Figure 5.9D,E). The erythroid and myeloid lineage trajectories were enriched for double mutant cells (Figure 5.9F).

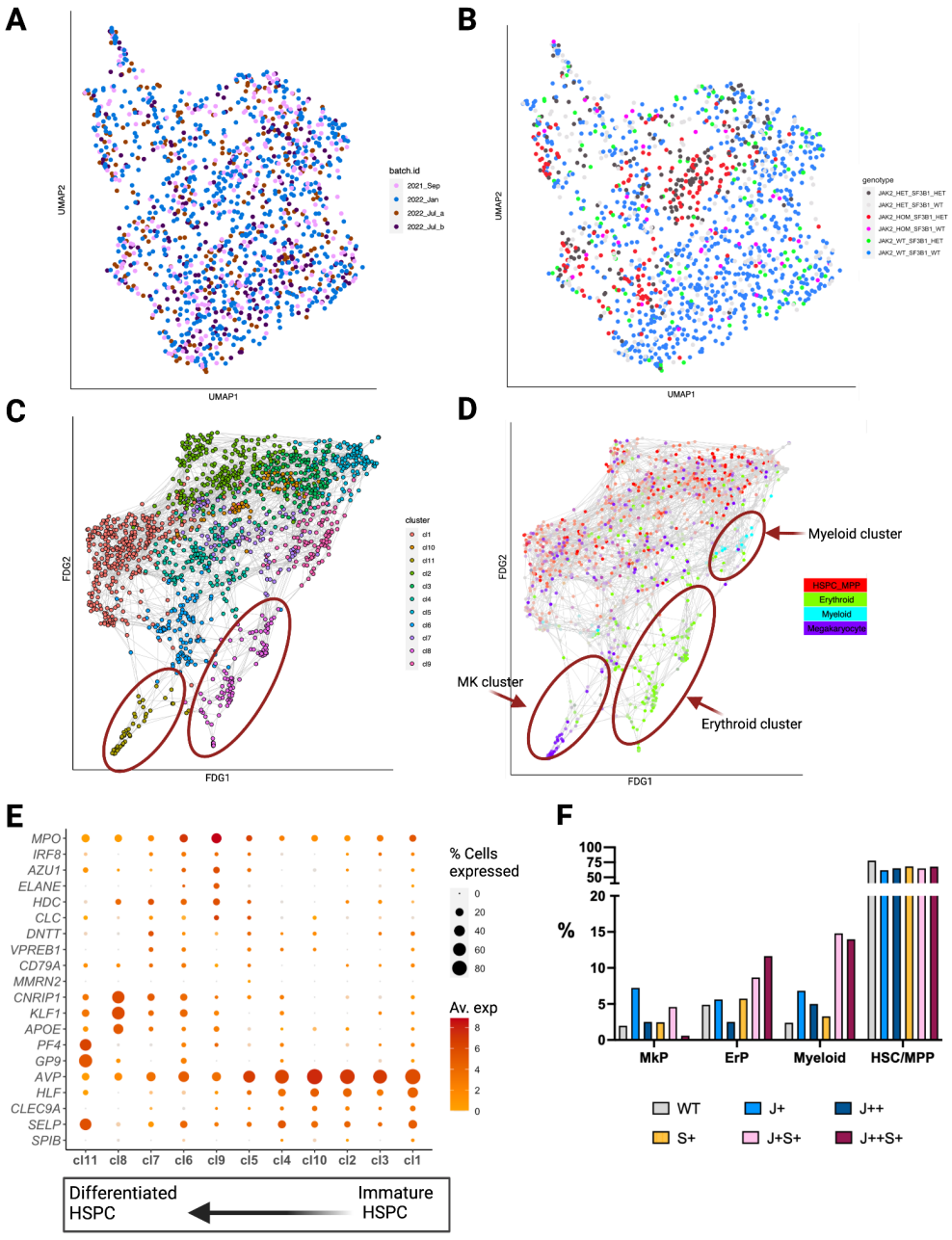


Figure 5.9. Analysis of TARGET-seq HSPC clusters. (A) UMAP plot of harmony integrated HSPCs coloured by sequencing batch. (B) UMAP plot of harmony integrated HSPCs coloured by genotype. (C) FDG plot of cell clusters. (D) FDG plot highlighting lineage signature genes. (E) Bubble plot of lineage

marker genes for each cluster. (F) Bar chart of genotype frequencies in MkP, ErP, myeloid and HSC/MPP clusters showing an enrichment of double mutant cells in ErP and myeloid clusters.

To further explore whether TARGET-seq HSPCs were enriched in specific progenitor populations, I projected TARGET-seq data onto the 10X HSPC gene expression data. Firstly, I separated HSPC populations from the 10X dataset, performed re-clustering and removed contaminating mature populations. Using a reference mapping approach[147], where similarities between two datasets are determined and the identified "anchors" are then transferred to direct integration of the datasets, I projected TARGET-seq HSPCs onto the reference 10X HSPC UMAP (Figure 5.10A). I derived predicted cell types from the shared anchors identified between the 10X and TARGET-seq datasets (Figure 5.10B). The majority of TARGET-seq HSPCs clustered within 10X HSPC clusters (75%) with no enrichment for any genotype in earlier HSC/MPP or later MEP-ERP/MKP populations (Figure 5.10C).

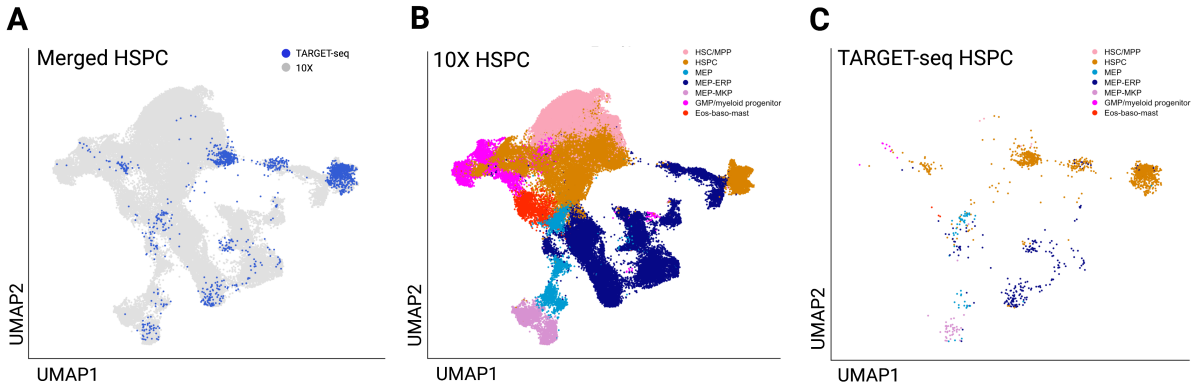


Figure 5.10. Projection of TARGET-seq onto 10X HSPC populations. (A) Merged HSPC UMAP plot with TARGET-seq cells highlighted in blue. UMAP plots of 10X HSPC (B) and TARGET-seq (C) populations overlaid onto the reference UMAP embeddings.

To examine the molecular signatures associated with *SF3B1*mut HSPCs, I performed differential gene expression analysis comparing *SF3B1*mut with *SF3B1*wt HSPCs to increase the power for detection of *SF3B1*mut -specific gene signatures. For this analysis, *SF3B1*mut HSPCs included both single- and double-mutant *SF3B1*mut HSPCs and *SF3B1*wt HSPCs comprised a mixture of *JAK2*V617Fmut and WT HSPCs across all patient and healthy control groups (Figure 5.11A). 605 genes were differentially expressed (DE) (144 upregulated and 461 downregulated genes), with 231 genes significantly DE (Figure 5.11B).

Mirroring the 10X HSPC analysis, I observed a downregulation of multiple pathways; TNFA signalling pathways, UV and inflammatory response and apoptosis pathways were downregulated in the *SF3B1*mut when compared with *SF3B1*wt HSPCs (Figure 5.11C). This pathway downregulation was specific to *SF3B1*mut when compared with *JAK2*V617Fmut HSPCs (Figure 5.11D). *SF3B1*mut showed upregulation in cell cycle pathways (E2F and MYC target pathways and G2M checkpoint pathways) when compared to all non-mutated WT HSPCs (Figure 5.11E). Tryptase genes were among the most highly DE genes which is analogous to the 10X HSPC analysis (Figure 5.11F). *IFI44L* and *CLEC11A* were also significantly increased in *SF3B1*mut HSPCs, again comparable to the 10X HSPC analysis. *PPP1R15A*, was downregulated in *SF3B1*mut HSPCs in the TARGET-seq data (and not in the 10X analysis).

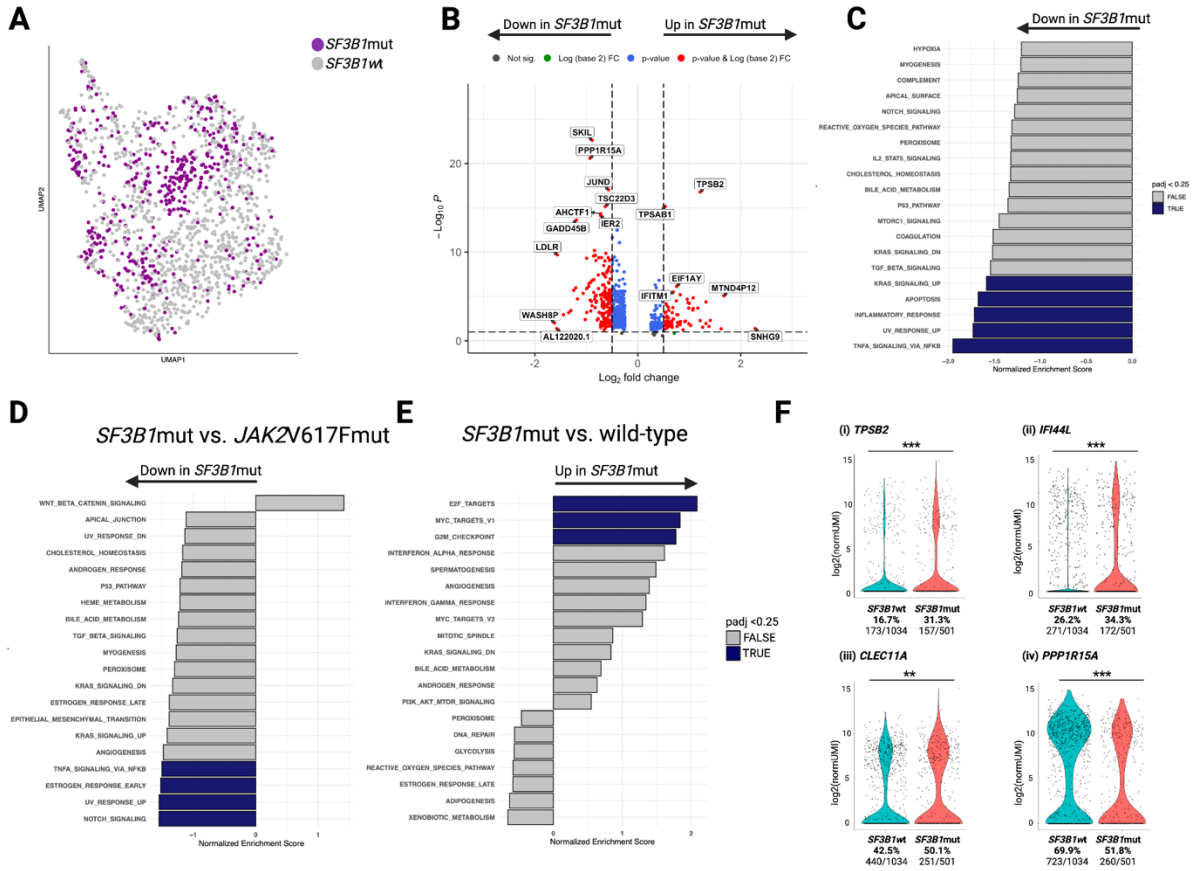


Figure 5.11. Differential gene expression analysis of TARGET-seq *SF3B1*mut vs *SF3B1*wt HSPCs. (A) UMAP plot showing *SF3B1*mut cells highlighted in purple. (B) Volcano plot highlighting top differentially expressed genes between *SF3B1*mut and *SF3B1*wt cells. Bar charts of GSEA results showing normalised enrichment scores (NES) and significant pathways (adjusted p-value <0.25) highlighted in navy comparing (C) *SF3B1*mut and *SF3B1*wt cells, (D) *SF3B1*mut and *JAK2*V617Fmut cells, (E) *SF3B1*mut and WT cells. (F) Violin plots of selected differentially expressed genes between *SF3B1*mut (red) and *SF3B1*wt (turquoise) HSPCs. Each dot representing the gene expression for each cell. Percentage of cells

expressing the gene presented below each plot. Wilcoxon test was used test statistical significance. NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

This gene codes for GADD34 protein, a protein phosphatase which is reported to dephosphorylate eIF-2-α (eukaryotic translation initiation factor 2 alpha subunit) which is part of unfolded protein response regulation[298]. Interestingly, this protein has been implicated in pulmonary fibrosis; *PPP1R15A* was reduced in patient lung fibroblasts, TGFβ-induced activation of fibroblasts decreased this further and pharmacological inhibition of PPP1R15A enhanced experimental models of lung fibrosis.[299] *SF3B1* has been shown to missplice another protein phosphatase, *PPP2R5A*, with reduced dephosphorylation of MYC and subsequent increased activation[289].

Importantly, TARGET-seq uniquely enabled intra-patient analysis where I was able to compare gene expression signatures between WT and mutant HSPCs in the same patient. I performed this in 3 patients where sufficient numbers of WT and mutant HSPCs were present to power the analysis; patient IF01144 had 31, 32 and 59 WT, *SF3B1* single mutant and double mutant HSPCs respectively and patient TNO123 had 87, 92 and 23 WT, *SF3B1* single mutant and *JAK2*V617F single mutant HSPCs respectively and patient IF01023 had 77 and 16, double-mutant and WT HSPCs, respectively for analysis (Figure 5.12A).

Intra-patient comparison *SF3B1*mut with *SF3B1*wt HSPCs showed some overlapping pathways across the patients but also dysregulated genes that were patient specific. For example, in patient IF011144, multiple pathways were enriched *SF3B1*mut HSPCs including adipogenesis, complement pathways and IL6/JAK/STAT3 signalling (Figure 5.12B). This aberrant signature was primarily driven by an *SF3B1*mut specific upregulation in erythroid and MK-related genes; *CD36*, *GATA1* and *ITGA2B* (Figure 5.12C). This was increased double as compared with single mutant HSPCs. Complement pathway signature genes such as *CD55* were also increased in *SF3B1*mut HSPCs (Figure 5.12C).

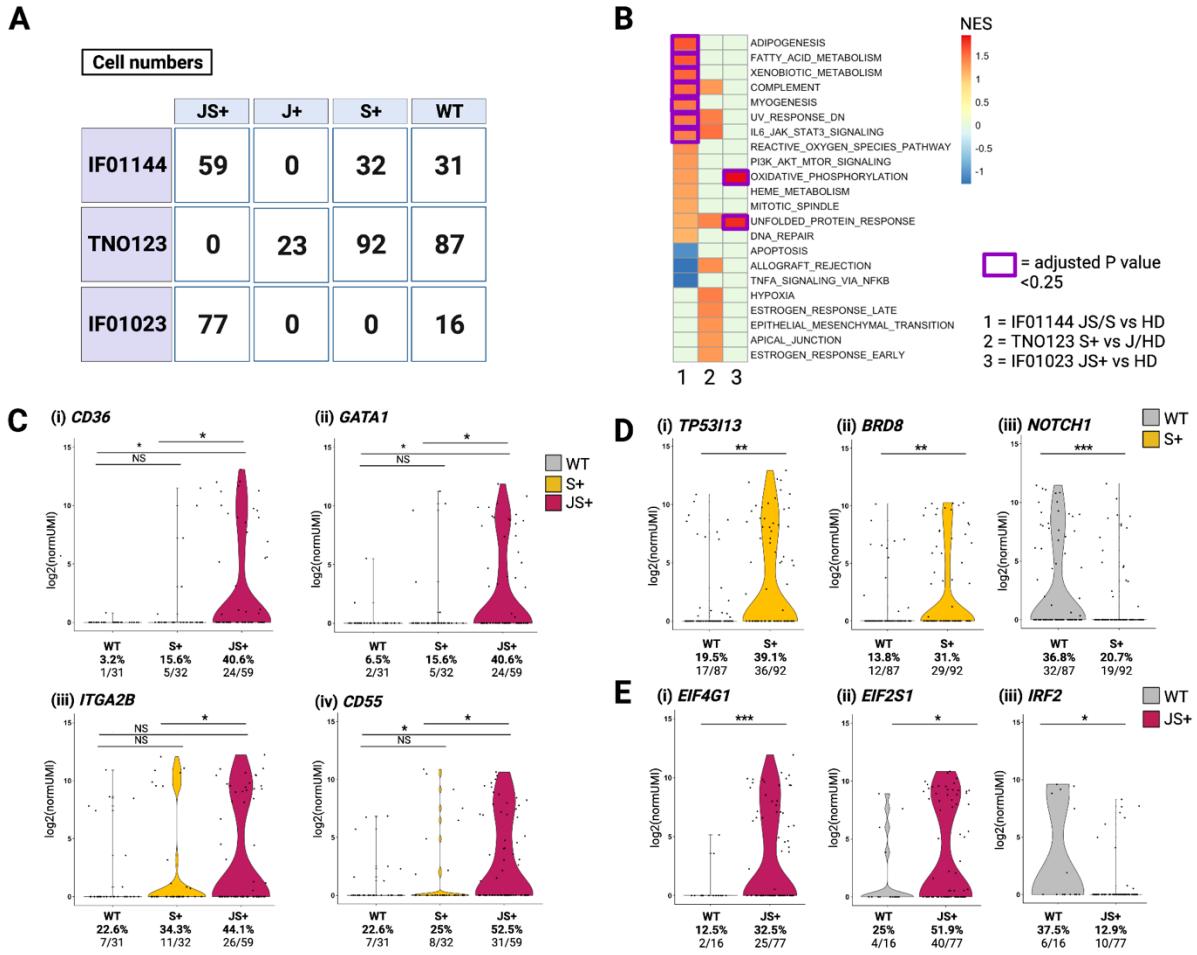


Figure 5.12. Differential gene expression analysis of TARGET-seq intra-patient mutant vs. WT HSPCs. (A) Cell numbers available of the 3 patients included in the analysis. (B) GSEA of Hallmark pathways comparing *SF3B1*mut with *SF3B1*wt cells for each patient. (C) Dysregulated genes in *SF3B1*^{mut} cells of patient IF01144; (i) *CD36*, (ii) *GATA1*, (iii) *ITGA2B*, (iv) *CD55*. (D) Dysregulated genes in *SF3B1*mut cells of patient TNO123; (i) *TP53I13*, (ii) *BRD8*, (iii) *NOTCH1*. (E) Dysregulated genes in *SF3B1*mut cells of patient IF01023; (i) *EIF4G1*, (ii) *EIF2S1*, (iii) *IRF2*. Each dot representing the gene expression for each cell. Percentage of cells expressing the gene presented below each plot. Wilcoxon test was used test statistical significance. NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

Patient TNO123 had single mutant HSPCs with *SF3B1*mut-specific dysregulated genes such as *TP53I13*, *BRD8* and *NOTCH1* (Figure 5.12D). Patient IF01023 in contrast had a majority of double mutant HSPCs which had distinct dysregulated signatures when compared to WT HSPCs; unfolded protein response genes were increased (*EIF4G1* and *EIF2S1*) and a reduced expression observed in genes such as *IRF2* in double mutant cells although worth considering low numbers of cells expressing this gene (Figure 5.12E).

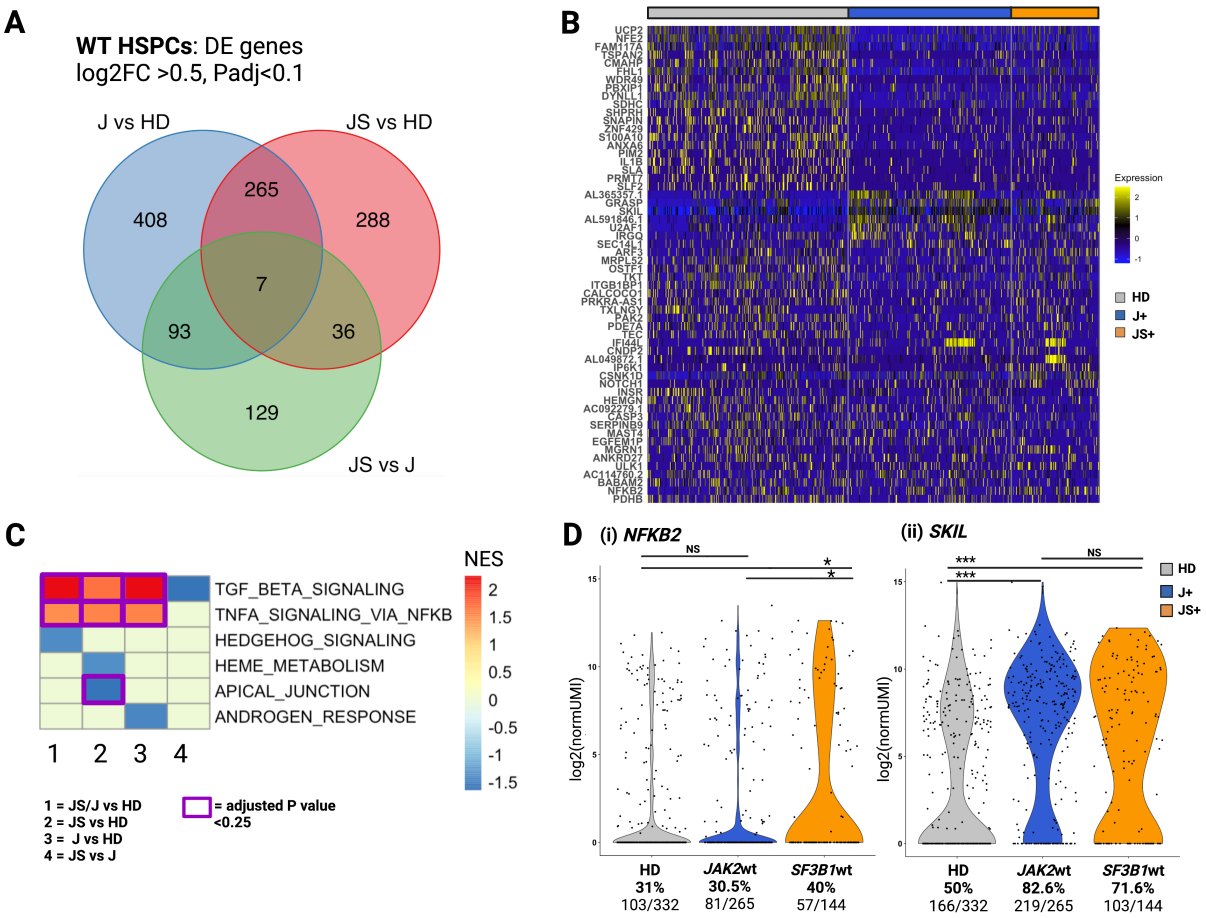


Figure 5.13. TARGET-seq WT HSPC analysis. (A) Venn diagram of differentially expressed genes. (B) Top differentially expressed genes between HD, J+ and JS+ WT HSPCs. (B) Heatmap of differentially expressed genes. (C) GSEA comparing WT HSPC of patient and HD. (D) Violin plots of select differentially expressed genes; (i) *NFKB2*, (ii) *SKIL*. Each dot representing the gene expression for each cell. Percentage of cells expressing the gene presented below each plotWilcoxon test was used test statistical significance. NS; not significant p-value ≥0.05, *p-value <0.05, **p-value <0.01, ***p-value <0.0001.

Finally, I performed an analysis focusing on WT HSPCs only comparing differential gene expression between patients in both genotype groups with healthy donors. In MPN, dysregulated signatures have been described in *JAK2*V617F patient WT HSPCs when compare with healthy WT HSPCs[122]. There were 741 cells included in this analysis; 332, 265 and 144 HSPCs from HD, J+ and JS+ patients respectively. There were 272 significantly DE genes shared between JS and J WT versus HD WT HSPCs (Figure 5.13A). Gene signatures between J+ and JS+ WT HSPCs were similar but distinct from WT HSPCs (Figure 5.13B). Pathway analysis was concordant with this showing an enrichment TGF beta signalling and TNFA signalling signatures in both J+ and JS+ WT HSPCs as compared with HD WT HSPCs (Figure 5.13C). Example genes in these pathways upregulated include *NFKB2* and *SKIL* genes (Figure 5.13D).

5-2.3 Aberrant splicing events in JS+ MPN

In a global analysis of splicing, more splicing events were detected in *JAK2*V617Fmut *SF3B1*mut doublet mutant HSPCs than *SF3B1*mut or *JAK2*V617Fmut single mutant HSPCs (Figure 5.14A).

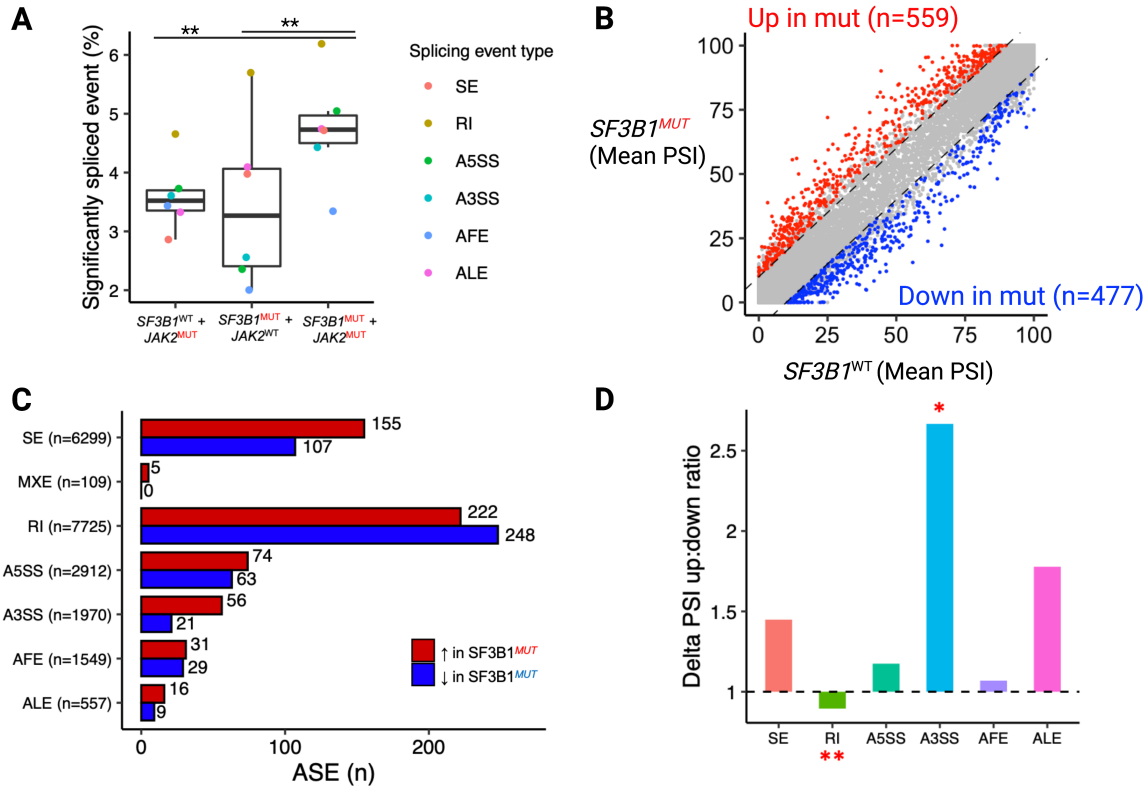


Figure 5.14. Global splicing events associated with *SF3B1*mut in the TARGET-seq dataset. (A) Boxplot showing the proportion of significantly spliced events in *JAK2*V617F only, *SF3B1*mut single mutant and *SF3B1*mut double mutant cells. (B) Scatterplot comparing mean percent spliced in (PSI) value of events between *SF3B1*mut (Y axis) to *SF3B1*wt cells (X axis); events significantly associated with *SF3B1*mut cells are coloured in red and those associated with *SF3B1*wt are coloured in blue. (C) Bar chart of splicing events increased (red) and decreased (blue) in *SF3B1*mut cells. (D) Bar chart showing delta of increased:decreased PSI (percent spliced in) ratio per event (*change in A3SS ratio highest, **decrease in RI). Bioinformatic analysis performed by Wei Wen, DPhil student in the Mead and Thongjuea laboratories.

More alternative splicing events (ASE) (n=559) were upregulated than downregulated (n=477) in *SF3B1*mut cells (Figure 5.14B). Retained intron events were the most prevalent ASE, followed by skipped exon (SE) events. A3SS events, the most often studied *SF3B1*mut -associated ASE, were enriched in *SF3B1*mut as compared with *SF3B1*wt cells (Figure 5.14C). A3SS were more often increased than decreased in *SF3B1*mut cells, consistent with the literature that *SF3B1*mut inserts cryptic A3SS[275,287,288,300,301] (Figure 5.14D). There was an increase in intron exclusion which was also consistent with the literature observed in *SF3B1*mut malignancies[226,285,302].

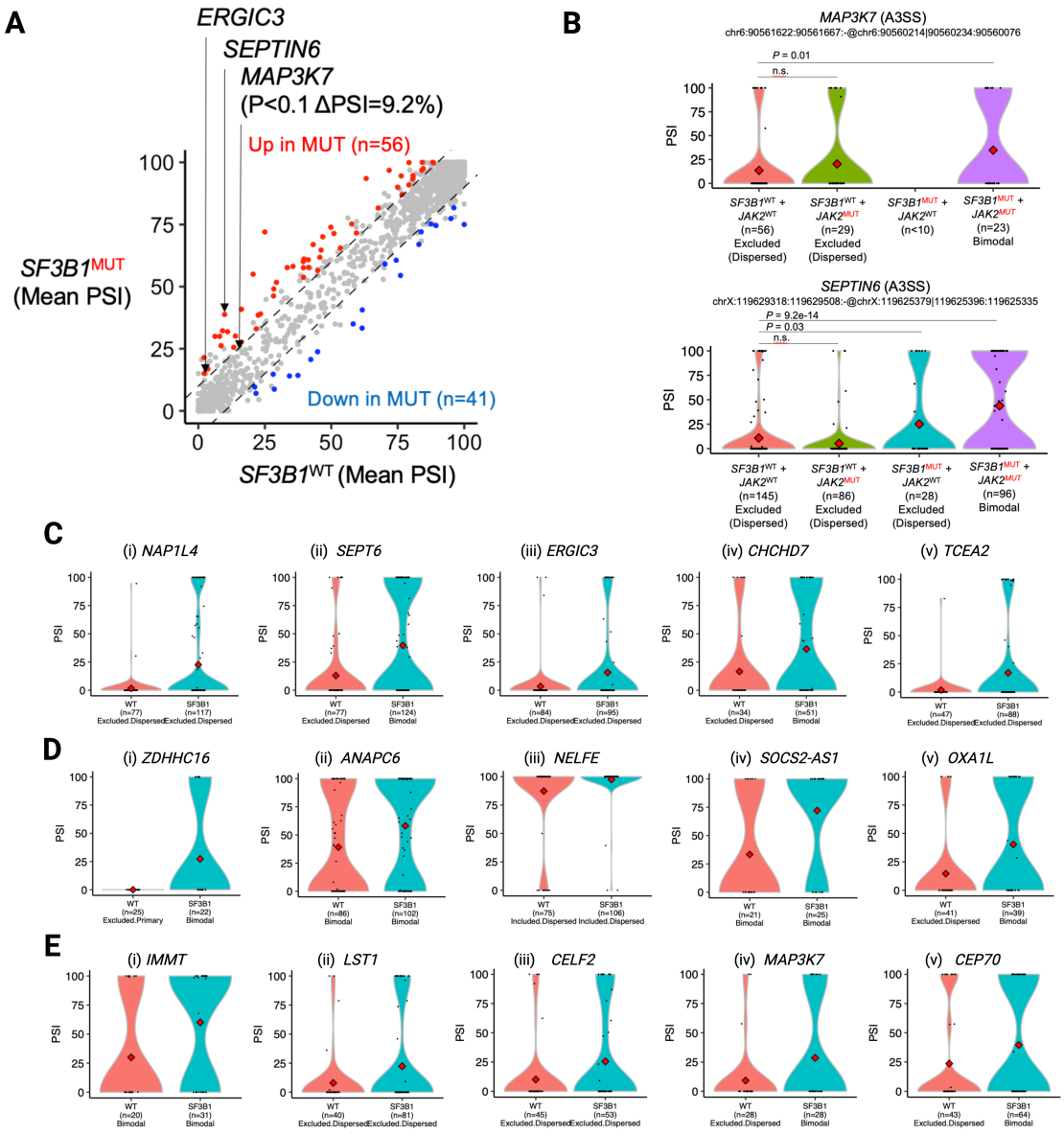
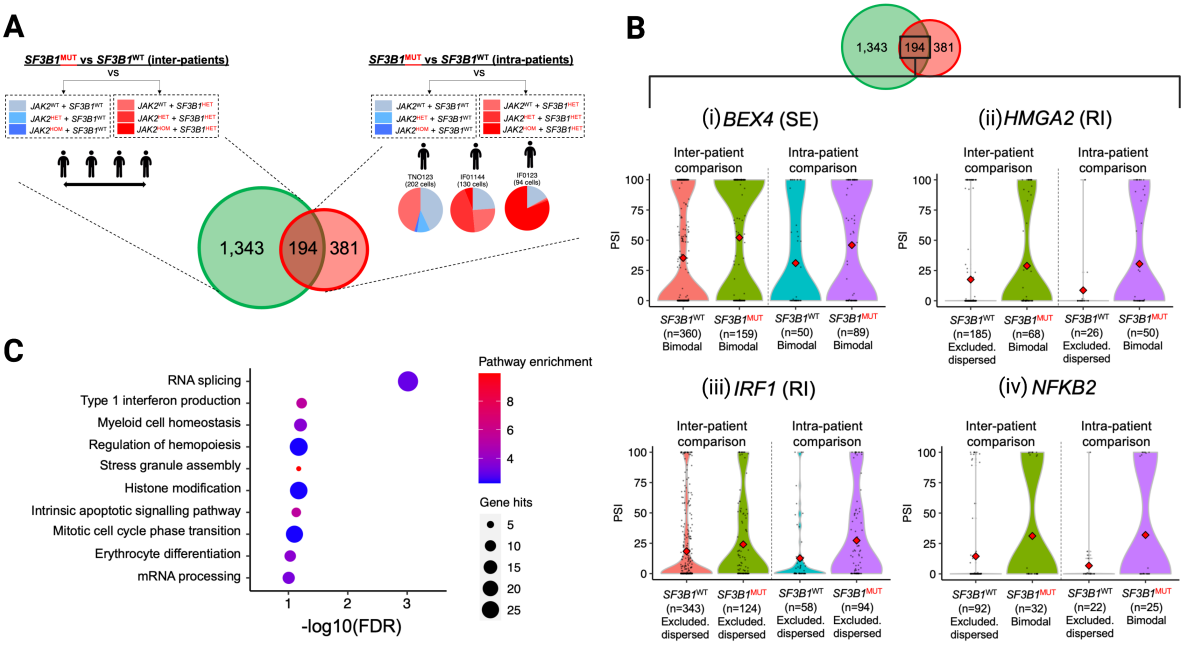


Figure 5.15. Identification of previously validated *SF3B1*mut-associated A3SS events. (A) Scatterplot showing splicing events *SF3B1*mut relative to *SF3B1*wt cells with significantly differentially spliced A3SS events coloured in red (increased in *SF3B1*mut) and blue (decreased in *SF3B1*mut). (B) Violin plots showing PSI of *MAP3K7* (upper plot) and *SEPTIN6* (lower plot) by genotype. (C-E) Violin plots of A3SS events validated in Beat AML cohort. Bioinformatic analysis performed by Wei Wen, DPhil student in the Mead and Thongjuea laboratories.

Next, I sought to identify previously validated *SF3B1*mut-associated ASE to determine validity of my dataset to interrogate for novel ASE. An A3SS in *MAP3K7*, the most widely studied in *SF3B1*mut MDS[225,273], was identified in *SF3B1*mut cells in addition to *ERGIC3*[127,226,286] and

SEPTIN6[286] (Figure 5.15A). *MAP3K7* and *SEPTIN6* A3SS events were more highly differentially spliced in double-mutant cells although fewer cells and therefore events were available for analysis in *SF3B1*mut single mutant cells (Figure 5.15B). Moreover, a comparison of patients with *SF3B1* K666 mutations in the Beat AML cohort[303] validated 15 *SF3B1*mut-associated ASEs (Figure5.14C-E).



D

Description	geneID
RNA splicing	TRA2A/CCDC130/CLK3/HNRNPA3/DDX39A/SNRNP200/DDX17/SF3A2/EIF4A3/DDX41/THOC6/PRPF40A/POLR2A/LUC7L/SLC38A2/FXR1/SRSF8/ZRSR2/PSIP1/ACIN1/WBP4/CELF2/WTAP/LUC7L3/DHX15/EFTUD2
Positive regulation of type I interferon production	NFKB2/STAT1/DDX41/CHUK/NFKB1/CTNNB1/POLR1C
Myeloid cell homeostasis	STAT1/RB1/HCLS1/ID2/JAK2/PIK3CB/ACIN1/NCKAP1L/IREB2/ANKRD54
Regulation of hemopoiesis	KMT2A/STAT1/TESPA1/PTK2B/PSMD13/ITGA2B/RB1/HCLS1/ID2/PSMC6/CLPTM1/MYB/CTNNB1/ADAM8/IL18/ACIN1/LMO2/NCKAP1L/MEIS1/ANKRD54
Intrinsic apoptotic signalling pathway in response to endoplasmic reticulum stress	DNAJC10/ATP2A3/ITPR1/HERPUD1/OPA1/MAP3K5
Regulation of mitotic cell cycle phase transition	PAFAH1B1/PSMG2/CTDSP1/NEDD1/PSMD13/RB1/ID2/PSMC6/CUL1/FHL1/ANAPC5/DDB1/APPL2/CUL3/CDC26/CEP57/DYNC1H1/ERCC3
Erythrocyte differentiation	STAT1/RB1/HCLS1/ID2/JAK2/ACIN1/NCKAP1L/ANKRD54

Figure 5.16. Differentially splicing events (DSE) in *SF3B1*mut HSPCs. (A) Venn diagram showing differentially spliced events derived from an inter-patient analysis comparing all *SF3B1*mut with *SF3B1*wt cells (green circle) and an intra-patient analysis comparing *SF3B1*mut with *SF3B1*wt cells (red circle) with the intersection the events common to both analyses. (B) Violin plots with select differentially spliced events common to both inter-patient and intra-patient analyses. (C) Pathway analysis of differentially spliced RI events between *SF3B1*mut and *SF3B1*wt HSPCs. (D) Leading edge analysis of genes in these top dysregulated pathways with events in genes relevant to JAK-STAT, NFKB signalling highlighted in red. Bioinformatic analysis performed by Wei Wen, DPhil student in the Mead and Thongjuea laboratories.

Akin to the TARGET-seq gene expression analysis approach, *SF3B1*mut cells were aggregated and compared with *SF3B1*wt cells, the latter inclusive of *JAK2*V617Fmut and WT cells. Then an intra-patient analysis was performed in patients where sufficient *SF3B1*mut cells were present for meaningful comparisons (Figure 5.16A-B).

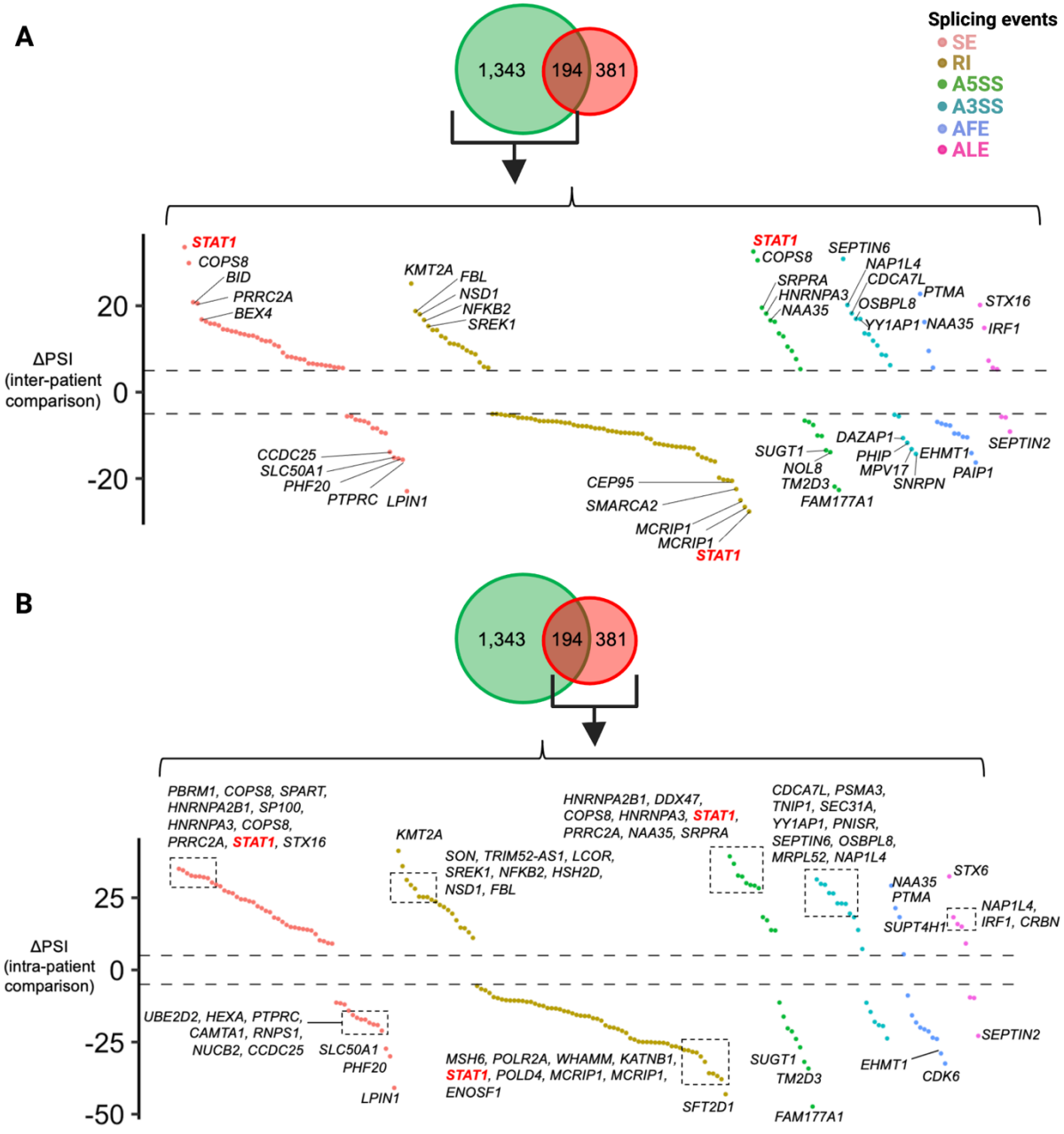


Figure 5.17. Top differentially spliced events comparing inter- and intra-patient results. Dot plots of the delta PSI (percent spliced in) of differentially spliced events detected in the inter-patient analysis (A) and intra-patient analysis (B). Events increased in *SF3B1*mut cells have a higher PSI and are higher on the Y-axis whereas events decreased in *SF3B1*mut cells have a negative delta PSA and are lower on the Y-axis. Events were coloured by event type; SE=skipped exon, RI=retained intron, A5SS=alternative 5' splice

site, A3SS=alternative 3' splice site, AFE=alternative first exon and ALE=alternative last exon. Bioinformatic analysis performed by Wei Wen, DPhil student in the Mead and Thongjuea laboratories.

Since RI events were the most commonly dysregulated ASEs, pathway enrichment analysis was performed of differentially spliced RI events (Figure 5.16C). For significantly DS reduced IR events, I observed enhancement in pathways expected in *SF3B1*mut myeloid neoplasms involving RNA splicing, regulation of translation, erythrocyte differentiation and inflammation response. Closer inspection of the genes driving dysregulation in these pathways highlighted relevant genes to the pathobiology to JS+ MPN (Figure 5.16D). ASE in *STAT1* gene were the among the top differentially spliced events in *SF3B1*mut cells in both inter- and intra-patient analyses (Figure 5.17A-B).

Two ASEs were detected in *STAT1*; one leading to reduced intron retention (RI) and a second skipped exon (SE) event. These events were mutually exclusive. Direct visualisation of these events was performed by merging cells by single-cell genotype to create a "pseudobulk", aligning these data to the reference genome and visualise the ASE using UCSC Genome Browser (Figure 5.18). The events were at the 3' end of *STAT1* and in close proximity. The SE event was a RI on direct visual inspection.

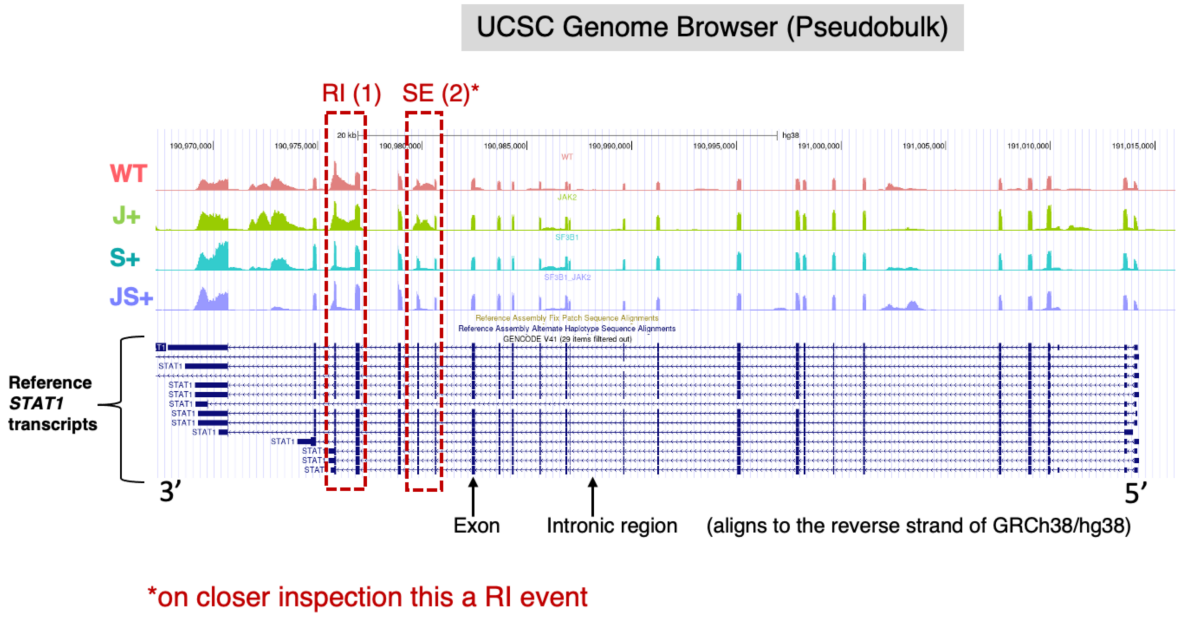


Figure 5.18. *STAT1* ASE visualisation. HSPCs were merged by single-cell genotype to create a "pseudobulk". The pseudobulk data was aligned to the reference genome (GRCh38/hg38) and the sequence visualised on UCSC Genome Browser. *STAT1* ASE are highlighted with red dashed boxes. Reference *STAT1* transcripts are plotted below the pseudobulk data. For the event "RI (1)", visualisation

was concordant with MARVEL results showing a reduction in intronic reads in S+ and JS+ pseudobulk data. The “SE (2)” event was in close proximity to RI (1) but on direct visualisation, this was a RI event rather than a SE with a reduction in intronic reads in S+ and JS+ pseudobulk data.

When an intron is included, it would be predicted to usually lead to NMD with a probable reduction in gene expression. The converse would be expected when an intron is excluded and therefore, RI reduction in *STAT1* would be predicted to cause increased gene activation and translation of *STAT1* and an increase in downstream *STAT1*-regulated pathways (Figure 5.19).

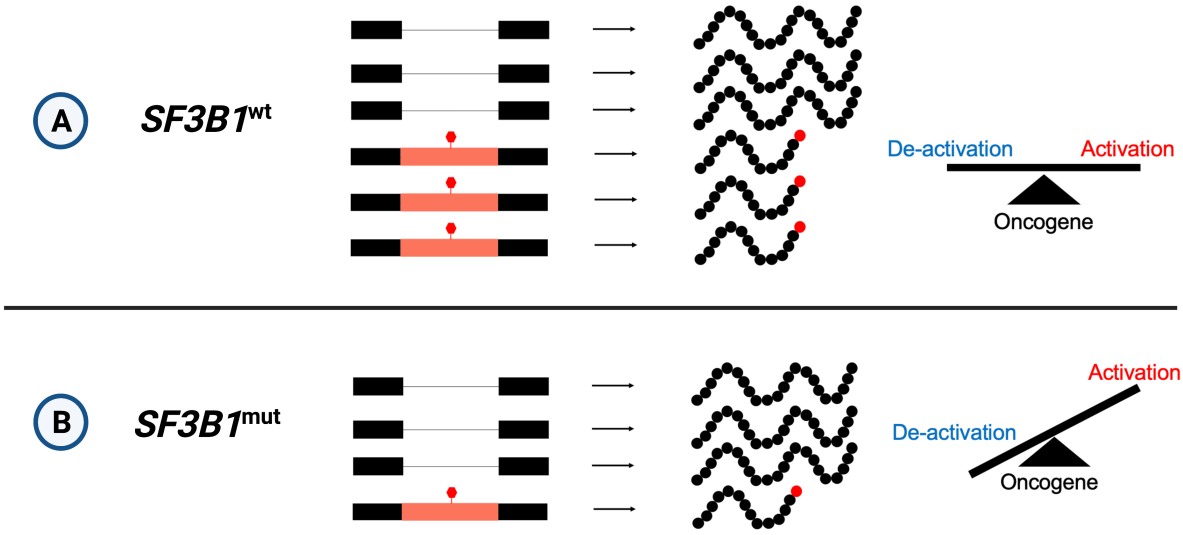


Figure 5.19. Schema of intron retention and effect on gene expression. (A) *SF3B1* WT scenario with intron retention and balance of gene activation. (B) *SF3B1* mutation with reduced intron retention and increased gene activation.

The *SF3B1*mut RI event, “RI (1)” was more significantly downregulated in *SF3B1*mut double mutant than *SF3B1*mut single mutant cells (Figure 5.20A). There was no significant difference in *STAT1* gene expression in TARGET-seq HSPCs (Figure 5.20B).

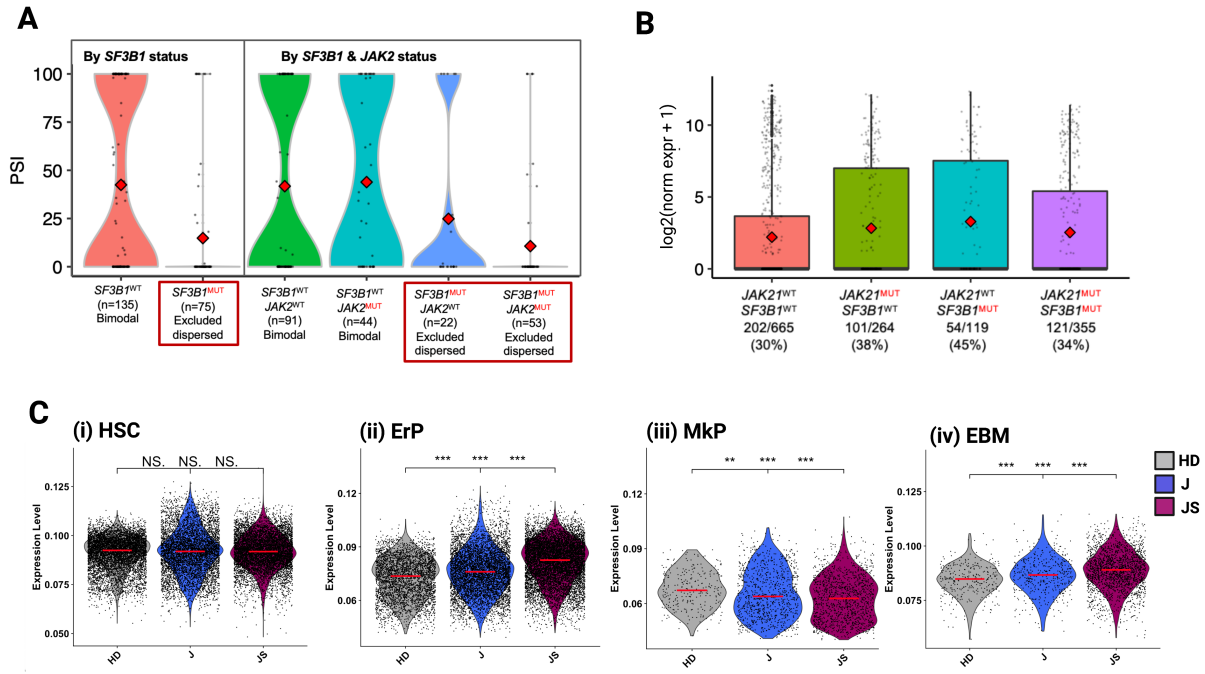


Figure 5.20. *STAT1* differential splicing and gene expression in TARGET-seq and 10X data. (A) PSI of RI event (1) by genotype in TARGET-seq data. (B) Box plot showing *STAT1* gene expression by genotype in TARGET-seq data. (C) *STAT1* target scores in specific HSPC populations shown in violin plots by donor group; (i) HSC, (ii) ErP, (iii) MkP and (iv) EBM populations. Bioinformatic analysis in (A) and (B) performed by Wei Wen, DPhil student in the Mead and Thongjuea laboratories.

To corroborate the hypothesis that aberrant splicing of *STAT1* leads to enhanced activation, I interrogated the 10X dataset for *STAT1* target genes using SCENIC regulon analysis and observed an upregulation in *STAT1* target gene scores in JS+ ErPs and EBM populations as compared with J+ and HD (Figure 5.20C).

5-3 Discussion

Applying a mutant-specific, single-cell multi-omic approach enabled me to uncover clonal composition of single HSPCs in *SF3B1*mut MPN patients. I observed acquisition of *SF3B1* before *JAK2V617F* in at least 2 patients which would not have been possible with a bulk sequencing approach. *SF3B1* mutations were detected throughout the haematopoietic progenitor tree without a specific genotype restricted or enriched within a particular HSPC compartment. However, there was an increase in *JAK2*V617F homozygous clones (single clones and those co-mutated with *SF3B1*) in CD34+CD38+ compartment. HSPCs carrying both mutations were enriched in

erythroid and myeloid differentiation trajectories as compared with cells that were single-mutant or WT. Increased *SF3B1* mutations were observed in a cell-specific manner, in abnormally expanded erythroid progenitors, in a related myeloid neoplasm, MDS, through application of the droplet-based multi-omic method, GoT-Splice[127].

TARGET-seq importantly facilitated detection of mutant-specific gene signatures showing that *SF3B1*mut gene signatures overlapped with the 10X HSPC dataset showing upregulation in tryptase genes, interferon-signalling genes and downregulation of *NFKB* signalling genes. Moreover, this method permitted analysis of WT patient HSPCs making intra-patient analyses and patient vs. HD WT HSPC analyses possible. Reproducing previous observations of non-mutant MPN HSPCs[122], MPN patient WT HSPCs exhibited increased inflammatory pathway signatures, particularly in genes involving TNFA and TGBF-beta signalling, consistent with suppression of normal haematopoiesis through cell-extrinsic inflammatory cues in MPN[205].

Mutant-specific single-cell splicing analysis was possible with full-length transcriptomic sequencing paired with high sensitivity genotyping and importantly low allelic dropout so accurate calling of WT calls was achieved. This is not reliably possible with other available methods including GoT-Splice[127], a new multi-omic single-cell platform which also aims to decipher mutant-specific splicing at the single-cell level. Although this technology incorporates long-read sequencing and is higher throughput, it has much lower genotyping sensitivity and therefore, is limited in confidently calling non-mutant cells. My analysis replicated previously validated events, such as in *MAP3K7* and *SEPTIN6*, but importantly uncovered novel *SF3B1*mut-associated ASE. Certain validated A3SS events were not identified in my dataset such as *ABCB7* due to insufficient gene expression in the HSPCs. Cryptic A3SS events are the most prevalent events described in association with *SF3B1* mutations[222,285-287]. A3SS were enriched in *SF3B1*mut cells in my data but RI events were more prevalent. Downregulation of RI events have been described in *SF3B1*mut malignancies (in MDS bulk short-read RNA sequencing[285] and CLL bulk Nanopore long-read sequencing)[226] but far less commonly than A3SS events. An explanation for this finding may be that *SF3B1*mut data has been predominantly derived using short-read sequencing methods which rely on in silico determination of ASE from exonic fragments and detection of transcripts with introns is challenging[304] often being miscategorised or not detected. Full-length sequencing methods, such as TARGET-seq, and long-read sequencing (Nanopore or PacBio)

can circumvent the limitations posed by short-read sequencing by covering longer regions and therefore interrogation of intron events is achievable.

Downregulation in intron retention in *STAT1* was one of the most highly differentially spliced events in *SF3B1*mut single and double mutant cells which has relevance to MPNs where JAK-STAT upregulation is central to the disease biology. The expected consequence of intron downregulation is increased *STAT1* activation. In support of this inference was the observation that downstream targets of *STAT1* were upregulated in abnormally expanded JS+ HSPC populations; ErP and EBM.

In MPNs, multiple STAT family proteins including STAT1 have been observed to be dysregulated. STAT1 is a transcription factor activated in response to interferon signalling triggered by various stimuli including microbes and is required for IFN-γ activation of quiescent HSCs[305]. STAT knockout (KO) murine models have illustrated the importance of these TFs in MPN biology. STAT5 is required for a *JAK2*V617F PV phenotype[46,306]. STAT3 KO murine models have been associated with divergent effects with thrombocytosis and promotion of a more severe phenotype in *JAK2*V617F murine model, although not required for the disease development[307,308]. In contrast, STAT3 KO in *MPL*W515L murine model conferred improved survival and a milder disease phenotype.[9] In STAT3 KO *JAK2*V617F mice, an increase in STAT1 expression and target genes was observed[308] which may explain the enhanced MPN phenotype. In colony assays from *JAK2*V617F PV and ET patients, *STAT1* was differentially expressed with increased *STAT1* in ET erythroblasts and upregulation of interferon signalling genes as compared with PV[45]. STAT1 KO in a *JAK2*V617F mouse model[309] was associated with increased erythropoiesis and reduced megakaryopoiesis. Based on this prior published data and the predicted effects of the STAT1 ASE in JS+, I would anticipate in JS+ MPN there is enhanced STAT1 signalling leading to an accelerated MPN phenotype.

Altogether, I have uncovered clonal composition of *SF3B1*mut MPN where the majority cells carry both mutations and importantly, *SF3B1*mut-associated aberrant transcriptomic and splicing changes are amplified when *SF3B1* co-exists with *JAK2*V617F in the same cell. *SF3B1*mut-associated ASE were identified that would be predicted to lead to increased JAK-STAT signalling and may explain the observed augmented MPN phenotype in *SF3B1*mut MPN.

Chapter 6: Discussion

MPNs are chronic pre-leukaemic malignancies with a long disease latency and few patients progress to leukaemia. This latency period, along with easily sampled tissue, denotes MPNs as excellent models for studying cancer evolution. MPNs are phenotypically heterogenous which reflects the genetic and molecular heterogeneity of the bone marrow compartment where the disease-initiating and -driving HSPCs originate.

An ideal MPN therapy would specifically target and eliminate disease-driving mutant HSPCs whilst leaving non-mutant cells unaffected and ultimately lead to disease remission and prevent progression. Current MPN therapies, especially JAK inhibitors such as RUX, have improved MF patient outcomes with spleen size reductions and symptom responses but evidence for disease modification has been limited[135]. Further, no MPN therapies have been shown to halt disease progression. Therefore, there is an evident need to understand mechanisms of disease evolution to develop effective therapies. Single-cell methodologies are ideally positioned to drive forward new understanding of MPN biology. These methods can specifically isolate mutant disease-driving rare HSPCs and interrogate their associate molecular signatures which is the fundamental requirement for identifying novel mutant specific biomarkers and therapeutic targets.

6-1 Outcomes of this research

Study of treatment-intolerant/resistant MPN patients uncovered distinct molecular profiles. HC-RES/INT ET patients had higher frequency of disease progression to MF than typically observed in ET enabling correlation of disease progression with mutation profiles. This analysis revealed an enrichment of *TP53* and *SF3B1* mutations in patients with myelofibrotic progression[187] which importantly was not alleviated with RUX treatment. Interestingly, *SF3B1* K666 hotspot predominated in my data and this *SF3B1* hotspot-disease phenotype association has been observed for example in MDS and uveal melanoma where K700 and R624 hotspots are enriched respectively. In a second more advanced MPN cohort of MF patients who were resistant or intolerant to RUX, I observed an enrichment in RAS-pathway mutations, an association which had not been reported previously[310].

In MDS, *SF3B1* mutations confer a milder disease phenotype in contrast to MPN which implies disease context and co-mutations (eg. JAK-STAT signalling mutations in MPN) are relevant. Understanding the HSPC clonal composition, haematopoietic cellular landscape and mutant-specific molecular and aberrant splicing signatures of JS+ MPN was critical to resolving mechanisms promoting accelerated fibrosis development. Delineation of the haematopoietic cellular landscape in *SF3B1*mut MPN was achieved through a high-throughput droplet multi-omic single-cell genomics approach (CITE-seq/10X genomics) which combines gene expression and proteomic (through CITE-seq), but not mutation, information per cell. This identified an expansion in earlier erythroid progenitors (ErP), megakaryocyte progenitors (MkP) and less expectedly EBM progenitors in JS+ HSPCs and in parallel, a reduction in lymphoid and monocyte cell populations. The expanded JS+ progenitor populations were associated with specific molecular signatures including genes conceivably relevant to fibrosis development such as tryptases and osteolectin. Interestingly, JS+ patients had a predominance of HSCs with upregulation in MK-lineage and IFN signalling genes. These HSCs were also enriched in tryptase genes suggesting that *SF3B1* acquisition may increase the proportion of "MK/EBM-primed" HSCs to promote enhanced aberrant differentiation of this MkP-ErP-EBM trajectory. Insights into how *SF3B1* aberrant splicing at the stem cell level may promote this abnormal differentiation may be addressed by applying TARGET-seq.

Deciphering the clonal architecture of JS+ MPN stem and progenitor cells was achieved by applying a second multi-omics genomics technique, TARGET-seq. This method pairs mutation information, gene expression (full-length transcriptome) and surface immunophenotype of the same single cell. High sensitivity mutation detection enables detection of mutant and WT cells and importantly, permits analysis of mutant-specific gene and splicing signatures. TARGET-seq applied to sorted single stem and progenitor cells demonstrated that in most cases *JAK2*V617F and *SF3B1* mutations typically (but not always) co-exist in the same cell. Furthermore, both *SF3B1* single mutant and JS+ double mutant cells were present in two patients indicating an *SF3B1* mutation was the initial event acquired and *JAK2*V617F was the secondary event. This reproduces previous single-cell analyses (TARGET-seq[122] and Fluidigm platform[297]) in *JAK2*V617F MPN observing that splicing factor mutations are more often initiating events. In MPN cases when other NDM are present, the JAK-STAT signalling driver mutation can occur as the antecedent event but also later as secondary event[85,297,311,312].

Gene signatures specific to TARGET-seq *SF3B1*mut HSPCs mirrored the signatures observed in 10X HSPCs such as upregulation of tryptase genes, interferon-signalling genes and downregulation in NF-κB pathway genes. Analysis of WT HSPCs highlighted enhanced inflammatory signatures of MPN patient (both J+ and JS+) HSPCs as compared with healthy controls. This reenforces prior analyses of non-mutant HSPCs displaying aberrant signatures[122,313,314] and the significance of the microenvironment in MPN disease pathogenesis.

Uniquely, full-length transcriptome TARGET-seq analysis made interrogation of *SF3B1*mut ASE possible uncovering novel events. *SF3B1*mut HSPCs had more retained intron events, in particular downregulation of retained introns. One of the top differentially spliced events was especially pertinent to dysregulated JAK-STAT signalling; a reduced intron event in *STAT1*. This event was predicted in silico to lead *SF3B1*mut reduced activation of *STAT1*. Upregulation of *STAT1* targets in abnormally expanded *SF3B1*mut HSPC populations in the 10X dataset supported this predicted *SF3B1*mut-associated increased *STAT1* activation. Upregulated STAT1 signalling has been demonstrated to be central to MPN disease biology. This analysis couples abnormal splicing with enhanced JAK-STAT signalling to support a mechanism for the clinical observation that *SF3B1* collaborates with *JAK2*V617F to promote an accelerated MPN phenotype.

6-2 Limitations of this research

Even though this analysis identified novel *SF3B1*mut-associated splicing events through full-length sequencing, relatively shorter fragments were sequenced (150bp length) which are aligned to the reference genome and ASE are determined *in silico*. Transcripts including intronic reads may not be detected and splicing events may be misclassified.

Secondly, while TARGET-seq detects mutations at high sensitivity, it is a lower to mid throughput method primarily because it is plate-based requiring multiple steps to generate libraries. Although this limits the numbers of cells that can be processed, automation of the library steps using liquid robotics platforms has improved processing capacity from hundreds to thousands of cells per week. This is relevant when discerning rare subclones which may be present at ~1-2% and can be difficult to distinguish a true clone from allelic dropout. However, this needs to be balanced with cost considerations for splicing analysis. The depth of sequencing required per cell to detect

most splicing events for downstream differential splicing analysis is reached at 1-2 million reads per cell and the cost may be prohibitive, restricting sequencing of sufficient cell numbers. Additionally, TARGET-seq transcriptomics is an adaption of smart-seq2 chemistry[315] which does not include UMIs which would improve resolution for confidently detecting single-cell isoforms. More recently, Smart-seq3[316] was developed which included a 5' UMI to overcome this limitation to improve single-cell isoform detection. Integrating this method with TARGET-seq may be advantageous more broadly for interrogating splicing factor mutated neoplasms.

Finally, the cohort included in the single-cell analyses consisted of MPN patients with a spectrum of clinical phenotypes and a mixture untreated and on various treatments. This was conducted to prioritise uniformity in genotypes in the cohort and to control for this, I ensured *JAK2*V617F controls were matched for gender, treatment, MPN subtype and where possible bulk level VAF and age. Tissue availability was also a factor in determining patients to include in the analysis considering *SF3B1* occur in ~10% of MF patients and <5% in earlier ET/PV. This heterogeneity may influence some of the molecular signatures detected. Another consideration in the patient selection was inclusion of *SF3B1*mut MDS cases as another comparator group but there is a significant breadth of published data on *SF3B1*mut MDS including single-cell analyses that could be re-analysed and compared with my dataset.

6-3 Future research

Validation and accurate annotation of the novel ASE identified in the TARGET-seq dataset is a next important step for this work. This would overcome the limitations posed by TARGET-seq full-length transcriptomics in accurately identifying ASE. This can be achieved through long-read sequencing methods which sequences the full length of a transcript. Nanopore technology[30] is a long-read sequencing approach that can capture the entire transcript length by passing a single DNA molecule through a nanopore membrane (or pore) under electric change and avoid PCR amplification. These experiments have been designed and will be performed shortly. A targeted approach is planned of the most highly differentially and biologically relevant ASE (eg. *STAT1*, *NFKB2*). This approach will resolve whether there are in fact multiple events in *STAT1* and sequencing the full length of the gene will establish if there are additional events. The experiments will be conducted on the same single cell cDNA generated from these TARGET-seq experiments.

Independent validation of *STAT1* ASE in MPN could be achieved through analysis of published RNA-sequencing datasets[293] inclusive of MDS cases. However, there are no available comparator single-cell MPN full-length sequencing datasets. Consequently, this method has utility if the ASE is identified but its absence isn't significant since bulk RNA-sequencing data may miss the event. Another method would be to generate cDNA (HSPC "mini-bulk"; eg. 100 cells) of additional *SF3B1*mut MPN cases and perform targeted sequencing of *STAT1*. Further validation may include confirmation of increased STAT1 expression at the protein level in JS+ MPN (eg. phosphorylated STAT1 by western blotting or IHC staining). Functional validation using genetic tools could be considered such as lentiviral vectors to overexpress the *STAT1* reduced intron transcript. Primary *JAK2*V617F CD34+ cells or *JAK2*V617F iPSCs could be transduced with the lentivirus expressing the transcript with liquid culture differentiation assays of erythroid, megakaryocyte and mast cell progenitor as potential read-outs.

The observation of abnormally expanded EBM progenitors in JS+ MPN as compared with J+ MPN has not been previously observed, although increased mast cells have been described in MF patient trephines[317,318] and murine MPN models[317]. Validation that this is more enhanced in JS+ MPN is underway using mast cell tryptase immunohistochemistry (IHC) staining of bone marrow trephines and comparing these with appropriate case controls (J+ cases). Correlation of mast cell frequency with fibrosis grade will be performed in parallel. Serial samples pre and post fibrosis transformation will be included where available which may elucidate whether mast cell expansion precedes fibrosis onset supporting the hypothesis that mast cells play a key role in MPN fibrosis development.

In this work, *JAK2*V617F and *SF3B1* mutations resided in separate clones in one case where. Conspicuously, this patient had no evidence of fibrosis. One hypothesis would be that both mutations are required in the same cell to mediate accelerated JAK-STAT signalling (e.g. STAT1) and enhance fibrosis development. However, additional cases would need to be examined to verify this.

A last but important consideration for splicing factor mutated neoplasms is aberrant splicing induced neoantigen generation and how these may be therapeutically harnessed. Cancer immunotherapies have rapidly evolved in the past decade transforming patient outcomes with the advent of cellular therapies, therapeutic antibodies and immune checkpoint inhibitor. A major

obstacle continues to be identification of tumour-specific or selective antigen targets. Neoantigens represent ideal such targets. These are proteins that are exclusively produced by cancer cells and not healthy cells, and towards which immune responses can be triggered. The commonest source are somatic mutations and in solid malignancies, neoantigen targeted immunotherapies are showing exciting results[319,320]. Alternative splicing is another mode of neoantigen generation and *SF3B1*-specific neoantigens have been identified in both MDS[321] and MPN[231]. The dataset generated in this research could be interrogated to identify potential alternative splicing induced neoantigens. Bioinformatic tools could be applied to computationally generate peptides from identified isoforms and computational algorithms such as NetMHCpan could be used to determine the MHC I & II binding affinity of these in silico generated peptides.

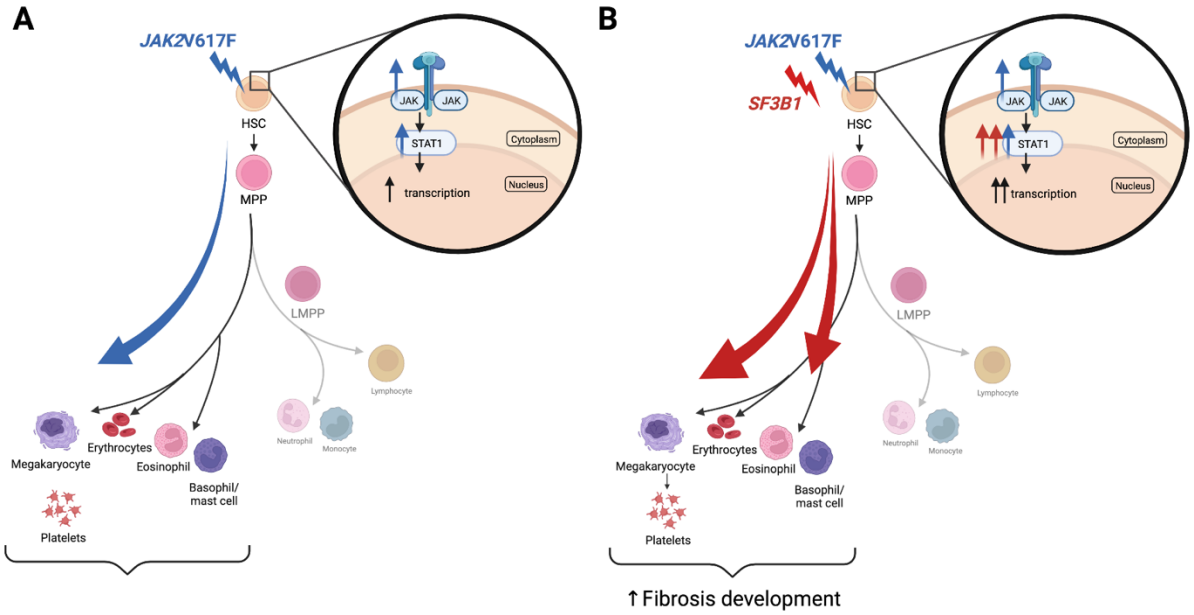


Figure 6. Proposed mechanism by which *SF3B1* mutations accelerate fibrosis development in *JAK2*V617F MPN. (A) *JAK2*V617F leads to increased JAK-STAT signalling and enhanced STAT1 activation promoting megakaryopoiesis. (B) When an *SF3B1* mutation is acquired –STAT1 is abberantly spliced leading to particularly enhanced STAT1 activation from the level of the HSC with marked upregulation of a shared differentiation trajectory with increased megakaryopoiesis but also basophil-mast cell development. These latter populations are implicated in fibrosis across many diseases and I hypothesis that these cells as well as MK may also play an important role in MPN fibrosis development.

6.1 Conclusion

The research presented in this thesis described the mutational landscape of high-risk treatment resistant/intolerant MPN cohorts. This analysis highlighted enhanced myelofibrotic disease progression in a specific subgroup, JS+ MPN. This observation generated the subsequent aims

of elucidating the cellular and molecular mechanisms promoting this progression. Applying a single-cell multi-omic approach successfully characterised the cellular and clonal architecture of JS+ MPN identifying abnormally expanded cell populations. Significantly, mutant-specific splicing signatures were also identified that would be expected to enhance JAK-STAT signalling and this supports a mechanism by which *SF3B1* cooperates with *JAK2*V617F to promote an accelerated MPN phenotype (Figure 6). The overarching aim is translating this new knowledge into patient benefit. Incorporation of spliceosome mutation information into clinical tools may improve risk stratification and specifically for JS+ MPN, a move towards precision medicine may be achievable through development of therapies targeting *STAT1*.

Bibliography

1. O'Sullivan J, Mead AJ. Heterogeneity in myeloproliferative neoplasms: Causes and consequences. *Adv Biol Regul*. Jan 2019;71:55-68. doi:10.1016/j.jbior.2018.11.007
2. O'Sullivan JM, Mead AJ, Psaila B. Single-cell methods in myeloproliferative neoplasms - old questions, new technologies. *Blood*. 2022;doi:10.1182/blood.2021014668
3. O'Sullivan JM, Hamblin A, Yap C, et al. The poor outcome in high molecular risk, hydroxycarbamide-resistant/intolerant ET is not ameliorated by ruxolitinib. *Blood*. Dec 5 2019;134(23):2107-2111. doi:10.1182/blood.2019001861
4. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. Jul 2022;36(7):1703-1719. doi:10.1038/s41375-022-01613-1
5. Barbui T, Tefferi A, Vannucchi AM, et al. Philadelphia chromosome-negative classical myeloproliferative neoplasms: revised management recommendations from European LeukemiaNet. *Leukemia*. 2018/05/01 2018;32(5):1057-1069. doi:10.1038/s41375-018-0077-1
6. Jamieson CH, Gotlib J, Durocher JA, et al. The JAK2 V617F mutation occurs in hematopoietic stem cells in polycythemia vera and predisposes toward erythroid differentiation. *Proc Natl Acad Sci U S A*. Apr 18 2006;103(16):6224-9. doi:10.1073/pnas.0601462103
7. Lundberg P, Takizawa H, Kubovcakova L, et al. Myeloproliferative neoplasms can be initiated from a single hematopoietic stem cell expressing JAK2-V617F. *J Exp Med*. Oct 20 2014;211(11):2213-30. doi:10.1084/jem.20131371
8. Rampal R, Al-Shahrour F, Abdel-Wahab O, et al. Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. *Blood*. May 29 2014;123(22):e123-33. doi:10.1182/blood-2014-02-554634
9. Kleppe M, Kwak M, Koppikar P, et al. JAK-STAT pathway activation in malignant and nonmalignant cells contributes to MPN pathogenesis and therapeutic response. *Cancer Discov*. Mar 2015;5(3):316-31. doi:10.1158/2159-8290.Cd-14-0736
10. Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet (London, England)*. Mar 19-25 2005;365(9464):1054-61. doi:10.1016/s0140-6736(05)71142-9
11. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med*. Apr 28 2005;352(17):1779-90. doi:10.1056/NEJMoa051113
12. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR Mutations in Myeloproliferative Neoplasms with Nonmutated JAK2. *New England Journal of Medicine*. 2013;369(25):2391-2405. doi:10.1056/NEJMoa1312542
13. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. *New England Journal of Medicine*. 2013;369(25):2379-2390. doi:10.1056/NEJMoa1311347
14. Pikman Y, Lee BH, Mercher T, et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med*. Jul 2006;3(7):e270. doi:10.1371/journal.pmed.0030270
15. Moulard O, Mehta J, Fryzek J, Olivares R, Iqbal U, Mesa RA. Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union. *Eur J Haematol*. Apr 2014;92(4):289-97. doi:10.1111/ejh.12256

16. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405. doi:10.1182/blood-2016-03-643544
17. Titmarsh GJ, Duncombe AS, McMullin MF, et al. How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am J Hematol*. Jun 2014;89(6):581-7. doi:10.1002/ajh.23690
18. Srour SA, Devesa SS, Morton LM, et al. Incidence and patient survival of myeloproliferative neoplasms and myelodysplastic/myeloproliferative neoplasms in the United States, 2001-12. *Br J Haematol*. Aug 2016;174(3):382-96. doi:10.1111/bjh.14061
19. Tefferi A, Betti S, Barraco D, et al. Gender and survival in essential thrombocythemia: A two-center study of 1,494 patients. *Am J Hematol*. Nov 2017;92(11):1193-1197. doi:10.1002/ajh.24882
20. Geyer HL, Kosiorek H, Dueck AC, et al. Associations between gender, disease features and symptom burden in patients with myeloproliferative neoplasms: an analysis by the MPN QOL International Working Group. *Haematologica*. Jan 2017;102(1):85-93. doi:10.3324/haematol.2016.149559
21. Tefferi A, Guglielmelli P, Larson DR, et al. Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood*. Oct 16 2014;124(16):2507-13; quiz 2615. doi:10.1182/blood-2014-05-579136
22. Harrison CN, Koschmieder S, Foltz L, et al. The impact of myeloproliferative neoplasms (MPNs) on patient quality of life and productivity: results from the international MPN Landmark survey. *Ann Hematol*. Oct 2017;96(10):1653-1665. doi:10.1007/s00277-017-3082-y
23. Dunbar AJ, Rampal RK, Levine R. Leukemia secondary to myeloproliferative neoplasms. *Blood*. Jul 2 2020;136(1):61-70. doi:10.1182/blood.2019000943
24. Psaila B, Mead AJ. Single-cell approaches reveal novel cellular pathways for megakaryocyte and erythroid differentiation. *Blood*. 2019;133(13):1427-1435. doi:10.1182/blood-2018-11-835371
25. Mead AJ, Mullally A. Myeloproliferative neoplasm stem cells. *Blood*. Mar 23 2017;129(12):1607-1616. doi:10.1182/blood-2016-10-696005
26. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *The New England journal of medicine*. Dec 19 2013;369(25):2391-2405. doi:10.1056/NEJMoa1312542
27. Delhommeau F, Dupont S, Tonetti C, et al. Evidence that the JAK2 G1849T (V617F) mutation occurs in a lymphomyeloid progenitor in polycythemia vera and idiopathic myelofibrosis. *Blood*. Jan 1 2007;109(1):71-7. doi:10.1182/blood-2006-03-007146
28. Angona A, Alvarez-Larran A, Bellosillo B, et al. Characterization of CD34+ hematopoietic progenitor cells in JAK2V617F and CALR-mutated myeloproliferative neoplasms. *Leuk Res*. Sep 2016;48:11-5. doi:10.1016/j.leukres.2016.06.009
29. Li J, Prins D, Park HJ, et al. Mutant calreticulin knockin mice develop thrombocytosis and myelofibrosis without a stem cell self-renewal advantage. 10.1182/blood-2017-09-806356. *Blood*. 2018;131(6):649.
30. Shide K, Kameda T, Yamaji T, et al. Calreticulin mutant mice develop essential thrombocythemia that is ameliorated by the JAK inhibitor ruxolitinib. *Leukemia*. May 2017;31(5):1136-1144. doi:10.1038/leu.2016.308

31. James C, Mazurier F, Dupont S, et al. The hematopoietic stem cell compartment of JAK2V617F-positive myeloproliferative disorders is a reflection of disease heterogeneity. 10.1182/blood-2008-02-137877. *Blood*. 2008;112(6):2429.
32. Kent DG, Li J, Tanna H, et al. Self-renewal of single mouse hematopoietic stem cells is reduced by JAK2V617F without compromising progenitor cell expansion. *PLoS Biol*. 2013;11(6):e1001576. doi:10.1371/journal.pbio.1001576
33. Mullally A, Lane SW, Ball B, et al. Physiological Jak2V617F expression causes a lethal myeloproliferative neoplasm with differential effects on hematopoietic stem and progenitor cells. *Cancer cell*. Jun 15 2010;17(6):584-96. doi:10.1016/j.ccr.2010.05.015
34. Scott LM, Tong W, Levine RL, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *The New England journal of medicine*. Feb 1 2007;356(5):459-68. doi:10.1056/NEJMoa065202
35. Pietra D, Rumi E, Ferretti VV, et al. Differential clinical effects of different mutation subtypes in CALR-mutant myeloproliferative neoplasms. *Leukemia*. Feb 2016;30(2):431-8. doi:10.1038/leu.2015.277
36. Elf S, Abdelfattah NS, Chen E, et al. Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. *Cancer Discov*. Apr 2016;6(4):368-81. doi:10.1158/2159-8290.Cd-15-1434
37. Chachoua I, Pecquet C, El-Khoury M, et al. Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood*. Mar 10 2016;127(10):1325-35. doi:10.1182/blood-2015-11-681932
38. Takei H, Morishita S, Araki M, et al. Detection of MPLW515L/K mutations and determination of allele frequencies with a single-tube PCR assay. *PLoS One*. 2014;9(8):e104958. doi:10.1371/journal.pone.0104958
39. Szuber N, Hanson CA, Lasho TL, et al. MPL-mutated essential thrombocythemia: a morphologic reappraisal. *Blood Cancer Journal*. 2018/11/20 2018;8(12):121. doi:10.1038/s41408-018-0159-3
40. Pardanani AD, Levine RL, Lasho T, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood*. Nov 15 2006;108(10):3472-6. doi:10.1182/blood-2006-04-018879
41. Haider M, Elala YC, Gangat N, Hanson CA, Tefferi A. MPL mutations and palpable splenomegaly are independent risk factors for fibrotic progression in essential thrombocythemia. *Blood cancer journal*. 2016;6(10):e487-e487. doi:10.1038/bcj.2016.98
42. Angona A, Fernández-Rodríguez C, Alvarez-Larrán A, et al. Molecular characterisation of triple negative essential thrombocythaemia patients by platelet analysis and targeted sequencing. Letter to the Editor. *Blood Cancer Journal*. 08/26/online 2016;6:e463. doi:10.1038/bcj.2016.75
43. Grinfeld J, Nangalia J, Baxter EJ, et al. Classification and Personalized Prognosis in Myeloproliferative Neoplasms. *New England Journal of Medicine*. 2018;379(15):1416-1430. doi:10.1056/NEJMoa1716614
44. Li B, Zhang L, Bai J, et al. Non-Driver Mutations Profile Identified By a 206-Gene NGS Panel in Patients with Primary Myelofibrosis and Post-Polycythaemic/Essential Thromocythaemia Myelofibrosis in a Single Center from China. *Blood*. 2016;128(22):1942.
45. Chen E, Beer PA, Godfrey AL, et al. Distinct clinical phenotypes associated with JAK2V617F reflect differential STAT1 signaling. *Cancer Cell*. Nov 16 2010;18(5):524-35. doi:10.1016/j.ccr.2010.10.013

46. Yan D, Hutchison RE, Mohi G. Critical requirement for Stat5 in a mouse model of polycythemia vera. *Blood*. Apr 12 2012;119(15):3539-49. doi:10.1182/blood-2011-03-345215
47. Walz C, Ahmed W, Lazarides K, et al. Essential role for Stat5a/b in myeloproliferative neoplasms induced by BCR-ABL1 and JAK2(V617F) in mice. *Blood*. Apr 12 2012;119(15):3550-60. doi:10.1182/blood-2011-12-397554
48. Lau WW, Hannah R, Green AR, Göttgens B. The JAK-STAT signaling pathway is differentially activated in CALR-positive compared with JAK2V617F-positive ET patients. *Blood*. Mar 5 2015;125(10):1679-81. doi:10.1182/blood-2014-12-618074
49. Kollmann K, Warsch W, Gonzalez-Arias C, et al. A novel signalling screen demonstrates that CALR mutations activate essential MAPK signalling and facilitate megakaryocyte differentiation. Original Article. *Leukemia*. 10/14/online 2016;31:934. doi:10.1038/leu.2016.280 https://www.nature.com/articles/leu2016280#supplementary-information
50. Dawson MA, Bannister AJ, Göttgens B, et al. JAK2 phosphorylates histone H3Y41 and excludes HP1α from chromatin. *Nature*. 09/27 2009;461(7265):819-822. doi:10.1038/nature08448
51. Liu F, Zhao X, Perna F, et al. JAK2V617F-mediated phosphorylation of PRMT5 downregulates its methyltransferase activity and promotes myeloproliferation. *Cancer cell*. Feb 15 2011;19(2):283-94. doi:10.1016/j.ccr.2010.12.020
52. Salati S, Genovese E, Carretta C, et al. Calreticulin Ins5 and Del52 mutations impair unfolded protein and oxidative stress responses in K562 cells expressing CALR mutants. *Scientific Reports*. 2019/07/22 2019;9(1):10558. doi:10.1038/s41598-019-46843-z
53. Ibarra J, Elbanna Y, Kurylowicz K, et al. Type 1 Calreticulin Mutations Differentially Activate the IRE1α-XBP1 Pathway of the Unfolded Protein Response to Drive Myeloproliferative Neoplasms. *Blood*. 2021/11/23/ 2021;138:628. doi:https://doi.org/10.1182/blood-2021-147859
54. Alvarez-Larrán A, Martínez-Avilés L, Hernández-Boluda JC, et al. Busulfan in patients with polycythemia vera or essential thrombocythemia refractory or intolerant to hydroxyurea. *Ann Hematol*. Dec 2014;93(12):2037-43. doi:10.1007/s00277-014-2152-7
55. Kiladjian JJ, Rain JD, Bernard JF, Briere J, Chomienne C, Fenaux P. Long-term incidence of hematological evolution in three French prospective studies of hydroxyurea and pipobroman in polycythemia vera and essential thrombocythemia. *Seminars in thrombosis and hemostasis*. Jun 2006;32(4 Pt 2):417-21. doi:10.1055/s-2006-942762
56. Harrison CN, Butt N, Campbell P, et al. Modification of British Committee for Standards in Haematology diagnostic criteria for essential thrombocythaemia. *British journal of haematology*. Nov 2014;167(3):421-3. doi:10.1111/bjh.12986
57. Kiladjian JJ, Cassinat B, Chevret S, et al. Pegylated interferon-alfa-2a induces complete hematologic and molecular responses with low toxicity in polycythemia vera. *Blood*. Oct 15 2008;112(8):3065-72. doi:10.1182/blood-2008-03-143537
58. Verger E, Soret-Dulphy J, Maslah N, et al. Ropeginterferon alpha-2b targets JAK2V617F-positive polycythemia vera cells in vitro and in vivo. *Blood cancer journal*. 2018;8(10):94-94. doi:10.1038/s41408-018-0133-0
59. Gisslinger H, Klade C, Georgiev P, et al. Final Results from PROUD-PV a Randomized Controlled Phase 3 Trial Comparing Ropeginterferon Alfa-2b to Hydroxyurea in Polycythemia Vera Patients. *Blood*. 2016;128(22):475.

60. Verger E, Cassinat B, Chauveau A, et al. Clinical and molecular response to interferon-alpha therapy in essential thrombocythemia patients with CALR mutations. *Blood*. Dec 10 2015;126(24):2585-91. doi:10.1182/blood-2015-07-659060
61. King KY, Matatall KA, Shen C-C, Goodell MA, Swierczek SI, Prchal JT. Comparative long-term effects of interferon α and hydroxyurea on human hematopoietic progenitor cells. *Experimental hematology*. 2015;43(10):912-918.e2. doi:10.1016/j.exphem.2015.05.013
62. Mullally A, Bruedigam C, Poveromo L, et al. Depletion of Jak2V617F myeloproliferative neoplasm-propagating stem cells by interferon-alpha in a murine model of polycythemia vera. *Blood*. May 2 2013;121(18):3692-702. doi:10.1182/blood-2012-05-432989
63. Mazur EM, Rosmarin AG, Sohl PA, Newton JL, Narendran A. Analysis of the mechanism of anagrelide-induced thrombocytopenia in humans. *Blood*. Apr 15 1992;79(8):1931-7.
64. Harrison CN, Campbell PJ, Buck G, et al. Hydroxyurea Compared with Anagrelide in High-Risk Essential Thrombocythemia. 2005;353(1):33-45. doi:10.1056/NEJMoa043800
65. Gisslinger H, Gotic M, Holowiecki J, et al. Final Results of the ANAHYDRET-Study: Non-Inferiority of Anagrelide Compared to Hydroxyurea in Newly Diagnosed WHO-Essential Thrombocythemia Patients. *Blood*. 2008;112(11):661.
66. Birgegård G, Besses C, Griesshammer M, et al. Treatment of essential thrombocythemia in Europe: a prospective long-term observational study of 3649 high-risk patients in the Evaluation of Anagrelide Efficacy and Long-term Safety study. *Haematologica*. 2018;103(1):51-60. doi:10.3324/haematol.2017.174672
67. Harrison C, Kiladjian JJ, Al-Ali HK, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *The New England journal of medicine*. Mar 1 2012;366(9):787-98. doi:10.1056/NEJMoa1110556
68. Verstovsek S, Vannucchi AM, Griesshammer M, et al. Ruxolitinib versus best available therapy in patients with polycythemia vera: 80-week follow-up from the RESPONSE trial. *Haematologica*. 2016;101(7):821-829. doi:10.3324/haematol.2016.143644
69. Harrison CN, Vannucchi AM, Kiladjian JJ, et al. Long-term findings from COMFORT-II, a phase 3 study of ruxolitinib vs best available therapy for myelofibrosis. *Leukemia*. 2016;30(8):1701-1707. doi:10.1038/leu.2016.148
70. Wang X, Ye F, Tripodi J, et al. JAK2 inhibitors do not affect stem cells present in the spleens of patients with myelofibrosis. *Blood*. Nov 6 2014;124(19):2987-95. doi:10.1182/blood-2014-02-558015
71. Raman I, Pasricha S-R, Prince HM, Yannakou CK. Management of hydroxyurea resistant or intolerant polycythemia vera. *Leukemia & Lymphoma*. 2021/08/24 2021;62(10):2310-2319. doi:10.1080/10428194.2021.1901092
72. Hernández-Boluda JC, Alvarez-Larrán A, Gómez M, et al. Clinical evaluation of the European LeukaemiaNet criteria for clinicohaematological response and resistance/intolerance to hydroxycarbamide in essential thrombocythaemia. *Br J Haematol*. Jan 2011;152(1):81-8. doi:10.1111/j.1365-2141.2010.08430.x
73. Harrison CN, Schaap N, Mesa RA. Management of myelofibrosis after ruxolitinib failure. *Ann Hematol*. Jun 2020;99(6):1177-1191. doi:10.1007/s00277-020-04002-9
74. Gupta V, Cerquozzi S, Foltz L, et al. Patterns of Ruxolitinib Therapy Failure and Its Management in Myelofibrosis: Perspectives of the Canadian Myeloproliferative Neoplasm Group. *JCO Oncology Practice*. 2020;16(7):351-359. doi:10.1200/jop.19.00506

75. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*. 2014;371(26):2488-2498. doi:10.1056/NEJMoa1408617
76. Cordua S, Kjaer L, Skov V, Pallisgaard N, Hasselbalch HC, Ellervik C. Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. *Blood*. 2019;134(5):469-479. doi:10.1182/blood.2019001113
77. Bowman RL, Busque L, Levine RL. Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. *Cell Stem Cell*. Feb 1 2018;22(2):157-170. doi:10.1016/j.stem.2018.01.011
78. Landgren O, Goldin LR, Kristinsson SY, Helgadottir EA, Samuelsson J, Björkholm M. Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24,577 first-degree relatives of 11,039 patients with myeloproliferative neoplasms in Sweden. *Blood*. Sep 15 2008;112(6):2199-204. doi:10.1182/blood-2008-03-143602
79. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nature genetics*. Apr 2009;41(4):446-9. doi:10.1038/ng.334
80. Soler G, Bernal-Vicente A, Antón AI, et al. The JAK2 46/1 haplotype does not predispose to CALR-mutated myeloproliferative neoplasms. *Ann Hematol*. May 2015;94(5):789-94. doi:10.1007/s00277-014-2266-y
81. Tapper W, Jones AV, Kralovics R, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat Commun*. Apr 7 2015;6:6691. doi:10.1038/ncomms7691
82. Hinds DA, Barnholt KE, Mesa RA, et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood*. Aug 25 2016;128(8):1121-8. doi:10.1182/blood-2015-06-652941
83. Bao EL, Nandakumar SK, Liao X, et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature*. 2020/10/01 2020;586(7831):769-775. doi:10.1038/s41586-020-2786-7
84. Alexandrov LB, Jones PH, Wedge DC, et al. Clock-like mutational processes in human somatic cells. *Nat Genet*. Dec 2015;47(12):1402-7. doi:10.1038/ng.3441
85. Williams N, Lee J, Mitchell E, et al. Life histories of myeloproliferative neoplasms inferred from phylogenies. *Nature*. 2022/02/01 2022;602(7895):162-168. doi:10.1038/s41586-021-04312-6
86. Van Egeren D, Escabi J, Nguyen M, et al. Reconstructing the Lineage Histories and Differentiation Trajectories of Individual Cancer Cells in Myeloproliferative Neoplasms. *Cell Stem Cell*. 2021;28(3):514-523.e9. doi:10.1016/j.stem.2021.02.001
87. Sousos N, Ni Leathlobhair M, Simoglou Karali C, et al. In utero origin of myelofibrosis presenting in adult monozygotic twins. *Nat Med*. Jun 2022;28(6):1207-1211. doi:10.1038/s41591-022-01793-4
88. Sanjuan-Pla A, Macaulay IC, Jensen CT, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*. 2013/10/01 2013;502(7470):232-236. doi:10.1038/nature12495
89. Carrelha J, Meng Y, Kettyle LM, et al. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*. 2018/02/01 2018;554(7690):106-111. doi:10.1038/nature25455

90. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature*. 2018/01/01 2018;553(7687):212-216. doi:10.1038/nature25168
91. Grover A, Sanjuan-Pla A, Thongjuea S, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun*. Mar 24 2016;7:11075. doi:10.1038/ncomms11075
92. Roch A, Trachsel V, Lutolf MP. Brief Report: Single-Cell Analysis Reveals Cell Division-Independent Emergence of Megakaryocytes From Phenotypic Hematopoietic Stem Cells. *Stem Cells*. Oct 2015;33(10):3152-7. doi:10.1002/stem.2106
93. Luis TC, Barkas N, Giustacchini A, et al. 2010 - Perivascular Niche Cells Sense Thrombocytopenia and Activate Platelet-Biased Hscs in an IL-1 Dependent Manner. *Experimental Hematology*. 2018/08/01/ 2018;64:S44. doi:https://doi.org/10.1016/j.exphem.2018.06.049
94. Ortmann CA, Kent DG, Nangalia J, et al. Effect of mutation order on myeloproliferative neoplasms. *The New England journal of medicine*. 2015;372(7):601-612. doi:10.1056/NEJMoa1412098
95. Nangalia J, Nice FL, Wedge DC, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica*. Nov 2015;100(11):e438-42. doi:10.3324/haematol.2015.129510
96. Usseglio F, Beaufils N, Calleja A, Raynaud S, Gabert J. Detection of CALR and MPL Mutations in Low Allelic Burden JAK2 V617F Essential Thrombocythemia. *J Mol Diagn*. Jan 2017;19(1):92-98. doi:10.1016/j.jmoldx.2016.08.006
97. Thompson ER, Nguyen T, Kankanige Y, et al. Clonal independence of JAK2 and CALR or MPL mutations in comutated myeloproliferative neoplasms demonstrated by single cell DNA sequencing. *Haematologica*. 08/13 2020;106(1):313-315. doi:10.3324/haematol.2020.260448
98. Curto-Garcia N, Harrison C, McLornan DP. Bone marrow niche dysregulation in myeloproliferative neoplasms. *Haematologica*. 2020;105(5):1189-1200. doi:10.3324/haematol.2019.243121
99. Schepers K, Pietras EM, Reynaud D, et al. Myeloproliferative neoplasia remodels the endosteal bone marrow niche into a self-reinforcing leukemic niche. *Cell Stem Cell*. Sep 5 2013;13(3):285-99. doi:10.1016/j.stem.2013.06.009
100. Kramann R, DiRocco DP, Humphreys BD. Understanding the origin, activation and regulation of matrix-producing myofibroblasts for treatment of fibrotic disease. *The Journal of Pathology*. 2013;231(3):273-289. doi:https://doi.org/10.1002/path.4253
101. El Agha E, Kramann R, Schneider RK, et al. Mesenchymal Stem Cells in Fibrotic Disease. *Cell Stem Cell*. Aug 3 2017;21(2):166-177. doi:10.1016/j.stem.2017.07.011
102. Di Buduo CA, Wray LS, Tozzi L, et al. Programmable 3D silk bone marrow niche for platelet generation ex vivo and modeling of megakaryopoiesis pathologies. *Blood*. Apr 2 2015;125(14):2254-64. doi:10.1182/blood-2014-08-595561
103. Pourcelot E, Trocme C, Mondet J, Bailly S, Toussaint B, Mossuz P. Cytokine profiles in polycythemia vera and essential thrombocythemia patients: clinical implications. *Exp Hematol*. May 2014;42(5):360-8. doi:10.1016/j.exphem.2014.01.006
104. Korn C RJ, Garcia-Garcia A, Fielding C, Khorshed R, Gonzalez-Anton S, Li Juan, Norfo R, Baxter JE, McKerrell T, Roberts T, Eldaly H, Godfrey AL, Castillo-Venzor A, Kusumbe A, Mead AJ, Green AR, Kent A, Lo Celso C, Mendez-Ferrer S. Niche Heterogeneity Impacts Evolution of Myeloproliferative Neoplasms Driven By the Same Oncogenic Pathway. Oral

presented at: ASH Annual Meeting 2018; 2018; San Diego, USA. Session Myeloproliferative Syndromes: Basic Science: Mechanism of Development and Progression
Hematology Disease Topics & Pathways.
105. Barbui T, Thiele J, Carobbio A, Vannucchi AM, Tefferi A. The rate of transformation from JAK2-mutated ET to PV is influenced by an accurate WHO-defined clinico-morphological diagnosis. Letter to the Editor. *Leukemia*. 11/26/online 2014;29:992. doi:10.1038/leu.2014.328
106. Li J, Kent DG, Godfrey AL, et al. JAK2V617F homozygosity drives a phenotypic switch in myeloproliferative neoplasms, but is insufficient to sustain disease. 10.1182/blood-2013-06-510222. *Blood*. 2014;123(20):3139.
107. Cerquozzi S, Tefferi A. Blast transformation and fibrotic progression in polycythemia vera and essential thrombocythemia: a literature review of incidence and risk factors. *Blood cancer journal*. 2015;5(11):e366-e366. doi:10.1038/bcj.2015.95
108. Alvarez-Larrán A, Cervantes F, Bellosillo B, et al. Essential thrombocythemia in young individuals: frequency and risk factors for vascular events and evolution to myelofibrosis in 126 patients. Original Article. *Leukemia*. 04/12/online 2007;21:1218. doi:10.1038/sj.leu.2404693
109. Tefferi A, Mudireddy M, Mannelli F, et al. Blast phase myeloproliferative neoplasm: Mayo-AGIMM study of 410 patients from two separate cohorts. *Leukemia*. May 2018;32(5):1200-1210. doi:10.1038/s41375-018-0019-y
110. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. Jun 9 2016;374(23):2209-2221. doi:10.1056/NEJMoa1516192
111. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat Med*. Oct 2020;26(10):1549-1556. doi:10.1038/s41591-020-1008-z
112. Rampal R, Ahn J, Abdel-Wahab O, et al. Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. 10.1073/pnas.1407792111. *Proceedings of the National Academy of Sciences*. 2014;111(50):E5401.
113. Luque Paz D, Jouanneau-Courville R, Riou J, et al. Leukemic evolution of polycythemia vera and essential thrombocythemia: genomic profiles predict time to transformation. *Blood Advances*. 2020;4(19):4887-4897. doi:10.1182/bloodadvances.2020002271
114. Kubesova B, Pavlova S, Malcikova J, et al. Low-burden TP53 mutations in chronic phase of myeloproliferative neoplasms: association with age, hydroxyurea administration, disease type and JAK2 mutational status. *Leukemia*. Feb 2018;32(2):450-461. doi:10.1038/leu.2017.230
115. Marcellino BK, Hoffman R, Tripodi J, et al. Advanced forms of MPNs are accompanied by chromosomal abnormalities that lead to dysregulation of TP53. *Blood Adv*. Dec 26 2018;2(24):3581-3589. doi:10.1182/bloodadvances.2018024018
116. Ramos CA, Bowman TA, Boles NC, et al. Evidence for diversity in transcriptional profiles of single hematopoietic stem cells. *PLoS Genet*. Sep 29 2006;2(9):e159. doi:10.1371/journal.pgen.0020159
117. Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Experimental & Molecular Medicine*. 2020/09/01 2020;52(9):1428-1442. doi:10.1038/s12276-020-0420-2
118. Ogbeide S, Giannese F, Mincarelli L, Macaulay IC. Into the multiverse: advances in single-cell multiomic profiling. *Trends in Genetics*. 2022;38(8):831-843. doi:10.1016/j.tig.2022.03.015

119. Kharchenko PV. The triumphs and limitations of computational methods for scRNA-seq. *Nature Methods*. 2021/07/01 2021;18(7):723-732. doi:10.1038/s41592-021-01171-x
120. Lareau CA, Ludwig LS, Muus C, et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nature Biotechnology*. 2021/04/01 2021;39(4):451-461. doi:10.1038/s41587-020-0645-6
121. Nam AS, Kim K-T, Chaligne R, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*. 2019/07/01 2019;571(7765):355-360. doi:10.1038/s41586-019-1367-0
122. Rodriguez-Meira A, Buck G, Clark SA, et al. Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol Cell*. Mar 21 2019;73(6):1292-1305.e8. doi:10.1016/j.molcel.2019.01.009
123. Rodriguez-Meira A, O'Sullivan J, Rahman H, Mead AJ. TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *STAR Protocols*. 2020/12/18/ 2020;1(3):100125. doi:https://doi.org/10.1016/j.xpro.2020.100125
124. Psaila B, Wang G, Rodriguez-Meira A, et al. Single-cell analyses reveal megakaryocyte-biased hematopoiesis in myelofibrosis and identify mutant clone-specific targets. *Molecular cell*. 2020;78(3):477-492. e8.
125. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021/06/24/ 2021;184(13):3573-3587.e29. doi:https://doi.org/10.1016/j.cell.2021.04.048
126. Myers RM, Izzo F, Kottapalli S, et al. Integrated Single-Cell Genotyping and Chromatin Accessibility Charts JAK2^{V617F} Human Hematopoietic Differentiation. *bioRxiv*. 2022:2022.05.11.491515. doi:10.1101/2022.05.11.491515
127. Gaiti F, Chamely P, Hawkins AG, et al. Single-cell multi-omics defines the cell-type specific impact of splicing aberrations in human hematopoietic clonal outgrowths. *bioRxiv*. 2022:2022.06.08.495292. doi:10.1101/2022.06.08.495292
128. Lebrigand K, Magnone V, Barbry P, Waldmann R. High throughput error corrected Nanopore single cell transcriptome sequencing. *Nature Communications*. 2020/08/12 2020;11(1):4025. doi:10.1038/s41467-020-17800-6
129. Durbin RM, Altshuler D, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010/10/01 2010;467(7319):1061-1073. doi:10.1038/nature09534
130. 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 Discovery*. 2019;9(8):1050-1063. doi:10.1158/2159-8290.Cd-18-1453
131. Wang F, Morita K, DiNardo CD, et al. Leukemia stemness and co-occurring mutations drive resistance to IDH inhibitors in acute myeloid leukemia. *Nature Communications*. 2021/05/10 2021;12(1):2607. doi:10.1038/s41467-021-22874-x
132. 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(4):536-551. doi:10.1158/2159-8290.CD-19-0710
133. Harrison C, Kiladjian J-J, Al-Ali HK, et al. JAK Inhibition with Ruxolitinib versus Best Available Therapy for Myelofibrosis. *New England Journal of Medicine*. 2012/03/01 2012;366(9):787-798. doi:10.1056/NEJMoa1110556

134. Verstovsek S, Mesa RA, Gotlib J, et al. A Double-Blind, Placebo-Controlled Trial of Ruxolitinib for Myelofibrosis. *New England Journal of Medicine*. 2012;366(9):799-807. doi:10.1056/NEJMoa1110557
135. Harrison CN, Vannucchi AM, Kiladjian JJ, et al. Long-term findings from COMFORT-II, a phase 3 study of ruxolitinib vs best available therapy for myelofibrosis. *Leukemia*. Aug 2016;30(8):1701-7. doi:10.1038/leu.2016.148
136. Harrison CN, Mead AJ, Panchal A, et al. Ruxolitinib vs best available therapy for ET intolerant or resistant to hydroxycarbamide. *Blood*. Oct 26 2017;130(17):1889-1897. doi:10.1182/blood-2017-05-785790
137. Barosi G, Besses C, Birgegard G, et al. A unified definition of clinical resistance/intolerance to hydroxyurea in essential thrombocythemia: results of a consensus process by an international working group. *Leukemia*. Feb 2007;21(2):277-80. doi:10.1038/sj.leu.2404473
138. Jones AV, Ward D, Lyon M, et al. Evaluation of methods to detect CALR mutations in myeloproliferative neoplasms. *Leukemia research*. Jan 2015;39(1):82-7. doi:10.1016/j.leukres.2014.11.019
139. Hamblin A, Burns A, Tham C, et al. Development and Evaluation of the Clinical Utility of a Next Generation Sequencing (NGS) Tool for Myeloid Disorders. *Blood*. 2014;124(21):2373.
140. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*. 2017/09/01 2017;14(9):865-868. doi:10.1038/nmeth.4380
141. Buus TB, Herrera A, Ivanova E, et al. Improving oligo-conjugated antibody signal in multimodal single-cell analysis. *Elife*. Apr 16 2021;10doi:10.7554/eLife.61973
142. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Controlled clinical trials*. Aug 1996;17(4):343-6.
143. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing V, Austria. Available online at https://www.R-project.org/.
144. Stoeckius M, Zheng S, Houck-Loomis B, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology*. 2018/12/19 2018;19(1):224. doi:10.1186/s13059-018-1603-1
145. Xi NM, Li JJ. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. *Cell Systems*. 2021/02/17/ 2021;12(2):176-194.e6. doi:https://doi.org/10.1016/j.cels.2020.11.008
146. Heaton H, Talman AM, Knights A, et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nature Methods*. 2020/06/01 2020;17(6):615-620. doi:10.1038/s41592-020-0820-1
147. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. Jun 13 2019;177(7):1888-1902.e21. doi:10.1016/j.cell.2019.05.031
148. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*. 2019/12/01 2019;16(12):1289-1296. doi:10.1038/s41592-019-0619-0
149. Wang G, Wen WX, Mead AJ, Roy A, Psaila B, Thongjuea S. Processing single-cell RNA-seq datasets using SingCellaR. *STAR Protoc*. Jun 17 2022;3(2):101266. doi:10.1016/j.xpro.2022.101266

150. Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nature Biotechnology*. 2022/02/01 2022;40(2):245-253. doi:10.1038/s41587-021-01033-z
151. Martin M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet Journal*. 2011;17:10-12.
152. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. Jan 1 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
153. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. Apr 1 2014;30(7):923-30. doi:10.1093/bioinformatics/btt656
154. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. Aug 15 2009;25(16):2078-9. doi:10.1093/bioinformatics/btp352
155. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. Dec 16 2019;20(1):278. doi:10.1186/s13059-019-1910-1
156. Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A*. Dec 23 2014;111(51):E5593-601. doi:10.1073/pnas.1419161111
157. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. Mar 15 2010;26(6):841-2. doi:10.1093/bioinformatics/btq033
158. Wen WX, Mead AJ, Thongjuea S. MARVEL: An integrated alternative splicing analysis platform for single-cell RNA sequencing data. *bioRxiv*. 2022:2022.08.25.505258. doi:10.1101/2022.08.25.505258
159. James C, Ugo V, Le Couedic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature*. Apr 28 2005;434(7037):1144-8. doi:10.1038/nature03546
160. Vannucchi AM, Kiladjian JJ, Griesshammer M, et al. Ruxolitinib versus standard therapy for the treatment of polycythemia vera. *N Engl J Med*. Jan 29 2015;372(5):426-35. doi:10.1056/NEJMoa1409002
161. Barosi G, Mesa R, Finazzi G, et al. Revised response criteria for polycythemia vera and essential thrombocythemia: an ELN and IWG-MRT consensus project. *Blood*. Jun 6 2013;121(23):4778-81. doi:10.1182/blood-2013-01-478891
162. Barosi G, Birgegard G, Finazzi G, et al. Response criteria for essential thrombocythemia and polycythemia vera: result of a European LeukemiaNet consensus conference. *Blood*. May 14 2009;113(20):4829-33. doi:10.1182/blood-2008-09-176818
163. Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. 10.1182/blood-2013-11-537167. *Blood*. 2014;123(14):2220.
164. Tefferi A, Lasho TL, Guglielmelli P, et al. Targeted deep sequencing in polycythemia vera and essential thrombocythemia. *Blood Adv*. Nov 29 2016;1(1):21-30. doi:10.1182/bloodadvances.2016000216
165. Asp J, Andréasson B, Hansson U, et al. Mutation status of essential thrombocythemia and primary myelofibrosis defines clinical outcome. *Haematologica*. 2016;101(4):e129. doi:10.3324/haematol.2015.138958
166. Vannucchi AM, Lasho TL, Guglielmelli P, et al. Mutations and prognosis in primary myelofibrosis. *Leukemia*. 2013/09/01 2013;27(9):1861-1869. doi:10.1038/leu.2013.119

167. Tefferi A, Guglielmelli P, Lasho TL, et al. MIPSS70+ Version 2.0: Mutation and Karyotype-Enhanced International Prognostic Scoring System for Primary Myelofibrosis. *Journal of Clinical Oncology*. 2018/06/10 2018;36(17):1769-1770. doi:10.1200/JCO.2018.78.9867
168. Santos FPS, Getta B, Masarova L, et al. Prognostic impact of RAS-pathway mutations in patients with myelofibrosis. *Leukemia*. 2020/03/01 2020;34(3):799-810. doi:10.1038/s41375-019-0603-9
169. Tefferi A, Lasho TL, Abdel-Wahab O, et al. IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. *Leukemia*. Jul 2010;24(7):1302-9. doi:10.1038/leu.2010.113
170. Patel KP, Newberry KJ, Luthra R, et al. Correlation of mutation profile and response in patients with myelofibrosis treated with ruxolitinib. *Blood*. 2015;126(6):790-797. doi:10.1182/blood-2015-03-633404
171. Spiegel JY, McNamara C, Kennedy JA, et al. Impact of genomic alterations on outcomes in myelofibrosis patients undergoing JAK1/2 inhibitor therapy. *Blood advances*. 2017;1(20):1729-1738. doi:10.1182/bloodadvances.2017009530
172. Pacilli A, Rotunno G, Mannarelli C, et al. Mutation landscape in patients with myelofibrosis receiving ruxolitinib or hydroxyurea. *Blood Cancer Journal*. 2018/11/22 2018;8(12):122. doi:10.1038/s41408-018-0152-x
173. Patnaik MM, Caramazza D, Gangat N, Hanson CA, Pardanani A, Tefferi A. Age and platelet count are IPSS-independent prognostic factors in young patients with primary myelofibrosis and complement IPSS in predicting very long or very short survival. *Eur J Haematol*. Feb 1 2010;84(2):105-8. doi:10.1111/j.1600-0609.2009.01373.x
174. Newberry KJ, Patel K, Masarova L, et al. Clonal evolution and outcomes in myelofibrosis after ruxolitinib discontinuation. *Blood*. Aug 31 2017;130(9):1125-1131. doi:10.1182/blood-2017-05-783225
175. Incyte Corp. Jakafi (ruxolitinib phosphate) [package insert]. U.S. Food and Drug Administration website. www.accessdata.fda.gov/drugsatfda_docs/label/2011/202192lbl.pdf. Revised November 2011. Accessed January 01, 2021.
176. Impact Biomedicines, Inc. Inrebic (fedratinib hydrochloride) [package insert]. U.S. Food and Drug Administration website. www.accessdata.fda.gov/drugsatfda_docs/label/2019/212327s000lbl.pdf#page=22. Revised August 2019. Accessed January 01, 2021.
177. Mesa RA, Vannucchi AM, Mead A, et al. Pacritinib versus best available therapy for the treatment of myelofibrosis irrespective of baseline cytopenias (PERSIST-1): an international, randomised, phase 3 trial. *Lancet Haematol*. May 2017;4(5):e225-e236. doi:10.1016/s2352-3026(17)30027-3
178. Mascarenhas J, Hoffman R, Talpaz M, et al. Pacritinib vs Best Available Therapy, Including Ruxolitinib, in Patients With Myelofibrosis: A Randomized Clinical Trial. *JAMA Oncology*. 2018;4(5):652-659. doi:10.1001/jamaoncol.2017.5818
179. Gerds AT, Savona MR, Scott BL, et al. Determining the recommended dose of pacritinib: results from the PAC203 dose-finding trial in advanced myelofibrosis. *Blood Advances*. 2020;4(22):5825-5835. doi:10.1182/bloodadvances.2020003314
180. Ortmann CA, Kent DG, Nangalia J, et al. Effect of Mutation Order on Myeloproliferative Neoplasms. *New England Journal of Medicine*. 2015;372(7):601-612. doi:10.1056/NEJMoa1412098

181. Tefferi A, Guglielmelli P, Lasho TL, et al. Mutation-Enhanced International Prognostic Systems for Essential Thrombocythemia (MIPSS-ET) and Polycythemia Vera (MIPSS-PV). *Blood*. 2018;132(Suppl 1):578. doi:10.1182/blood-2018-99-109715
182. Chowdhury O, O'Sullivan J, Barkas N, et al. Spliceosome mutations are common in persons with myeloproliferative neoplasm-associated myelofibrosis with RBC-transfusion-dependence and correlate with response to pomalidomide. *Leukemia*. 2020/08/07 2020;doi:10.1038/s41375-020-0979-6
183. Brecqueville M, Rey J, Devillier R, et al. Array comparative genomic hybridization and sequencing of 23 genes in 80 patients with myelofibrosis at chronic or acute phase. *Haematologica*. Jan 2014;99(1):37-45. doi:10.3324/haematol.2013.091454
184. Coltro G, Rotunno G, Mannelli L, et al. RAS/CBL mutations predict resistance to JAK inhibitors in myelofibrosis and are associated with poor prognostic features. *Blood Advances*. 2020;4(15):3677-3687. doi:10.1182/bloodadvances.2020002175
185. Guglielmelli P, Biamonte F, Rotunno G, et al. Impact of mutational status on outcomes in myelofibrosis patients treated with ruxolitinib in the COMFORT-II study. *Blood*. 2014;123(14):2157-2160. doi:10.1182/blood-2013-11-536557
186. Verstovsek S, Hasserjian RP, Pozdnyakova O, et al. PRM-151 in Myelofibrosis: Efficacy and Safety in an Open Label Extension Study. *Blood*. 2018/11/29/ 2018;132:686. doi:https://doi.org/10.1182/blood-2018-99-115362
187. O'Sullivan JM, Hamblin A, Yap C, et al. The poor outcome in high molecular risk, hydroxycarbamide-resistant/intolerant ET is not ameliorated by ruxolitinib. *Blood*. 2019;134(23):2107-2111. doi:10.1182/blood.2019001861
188. Tefferi A, Lasho TL, Guglielmelli P, et al. Targeted deep sequencing in polycythemia vera and essential thrombocythemia. *Blood Advances*. 2016;1(1):21. doi:10.1182/bloodadvances.2016000216
189. Kubesova B, Pavlova S, Malcikova J, et al. Low-burden TP53 mutations in chronic phase of myeloproliferative neoplasms: association with age, hydroxyurea administration, disease type and JAK2 mutational status. Original Article. *Leukemia*. 07/24/online 2017;32:450. doi:10.1038/leu.2017.230
https://www.nature.com/articles/leu2017230#supplementary-information
190. Beer PA, Delhommeau F, LeCouédic J-P, et al. Two routes to leukemic transformation after a JAK2 mutation–positive myeloproliferative neoplasm. *Blood*. 2010;115(14):2891. doi:10.1182/blood-2009-08-236596
191. Senín A, Fernández-Rodríguez C, Bellosillo B, et al. Non-driver mutations in patients with JAK2V617F-mutated polycythemia vera or essential thrombocythemia with long-term molecular follow-up. *Annals of Hematology*. 2018/03/01 2018;97(3):443-451. doi:10.1007/s00277-017-3193-5
192. Tefferi A, Guglielmelli P, Lasho TL, et al. Mutation-enhanced international prognostic systems for essential thrombocythaemia and polycythaemia vera. *Br J Haematol*. Apr 2020;189(2):291-302. doi:10.1111/bjh.16380
193. Guglielmelli P, Biamonte F, Rotunno G, et al. Impact of mutational status on outcomes in myelofibrosis patients treated with ruxolitinib in the COMFORT-II study. *Blood*. 2014;123(14):2157. doi:10.1182/blood-2013-11-536557
194. Harrison C, Nangalia J, Boucher RH, et al. Ruxolitinib Versus Best Available Therapy for PV Intolerant or Resistant to Hydroxycarbamide in a Randomized Trial. *Blood*. 2022;140(Supplement 1):1781-1783. doi:10.1182/blood-2022-157273

195. Malcovati L, Papaemmanuil E, Bowen DT, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood*. 2011;118(24):6239. doi:10.1182/blood-2011-09-377275
196. Malcovati L, Karimi M, Papaemmanuil E, et al. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood*. Jul 9 2015;126(2):233-41. doi:10.1182/blood-2015-03-633537
197. Mian SA, Smith AE, Kulasekararaj AG, et al. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. *Haematologica*. 2013;98(7):1058-1066. doi:10.3324/haematol.2012.075325
198. Broseus J, Alpermann T, Wulfert M, et al. Age, JAK2(V617F) and SF3B1 mutations are the main predicting factors for survival in refractory anaemia with ring sideroblasts and marked thrombocytosis. *Leukemia*. Sep 2013;27(9):1826-31. doi:10.1038/leu.2013.120
199. Makishima H, Visconte V, Sakaguchi H, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. 2012;119(14):3203. doi:10.1182/blood-2011-12-399774
200. Patnaik MM, Lasho TL, Hodnefield JM, et al. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. *Blood*. 2012;119(2):569. doi:10.1182/blood-2011-09-377994
201. Kanagal-Shamanna R, Montalban-Bravo G, Sasaki K, et al. Only SF3B1 mutation involving K700E independently predicts overall survival in myelodysplastic syndromes. *Cancer*. Oct 1 2021;127(19):3552-3565. doi:10.1002/cncr.33745
202. Dalton WB, Helmenstine E, Pieterse L, et al. The K666N mutation in SF3B1 is associated with increased progression of MDS and distinct RNA splicing. *Blood Advances*. 2020;4(7):1192-1196. doi:10.1182/bloodadvances.2019001127
203. Harbour JW, Roberson ED, Anbunathan H, Onken MD, Worley LA, Bowcock AM. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat Genet*. Feb 2013;45(2):133-5. doi:10.1038/ng.2523
204. Loscocco GG, Guglielmelli P, Mannelli F, et al. SF3B1 mutations in primary and secondary myelofibrosis: Clinical, molecular and prognostic correlates. *American Journal of Hematology*. 2022;97(9):E347-E349. doi:https://doi.org/10.1002/ajh.26648
205. Fisher DAC, Fowles JS, Zhou A, Oh ST. Inflammatory Pathophysiology as a Contributor to Myeloproliferative Neoplasms. *Front Immunol*. 2021;12:683401. doi:10.3389/fimmu.2021.683401
206. Mylonas E, Yoshida K, Frick M, et al. Single-cell analysis based dissection of clonality in myelofibrosis. *Nat Commun*. Jan 7 2020;11(1):73. doi:10.1038/s41467-019-13892-x
207. 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*. Aug 2019;9(8):1050-1063. doi:10.1158/2159-8290.Cd-18-1453
208. Maslah N, Roux B, Kaci N, et al. JAK Inhibition Mediates Clonal Selection of RAS Pathway Mutations in Myeloproliferative Neoplasms. *Blood*. 2022;140(Supplement 1):795-796. doi:10.1182/blood-2022-167719
209. Stivala S, Codilupi T, Brkic S, et al. Targeting compensatory MEK/ERK activation increases JAK inhibitor efficacy in myeloproliferative neoplasms. *J Clin Invest*. 2019;129(4):1596-1611. doi:10.1172/jci98785 Accessed 2019/03//. http://europepmc.org/abstract/MED/30730307

https://doi.org/10.1172/JCI98785
https://europepmc.org/articles/PMC6436863
https://europepmc.org/articles/PMC6436863?pdf=render
210. Stanley RF, Abdel-Wahab O. Dysregulation and therapeutic targeting of RNA splicing in cancer. *Nature Cancer*. 2022/05/01 2022;3(5):536-546. doi:10.1038/s43018-022-00384-z
211. Zhou Z, Gong Q, Wang Y, et al. The biological function and clinical significance of SF3B1 mutations in cancer. *Biomarker Research*. 2020/09/03 2020;8(1):38. doi:10.1186/s40364-020-00220-5
212. Kahles A, Lehmann KV, Toussaint NC, et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer cell*. Aug 13 2018;34(2):211-224.e6. doi:10.1016/j.ccell.2018.07.001
213. Seiler M, Peng S, Agrawal AA, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Reports*. 2018/04/03/ 2018;23(1):282-296.e4. doi:https://doi.org/10.1016/j.celrep.2018.01.088
214. Lee SC-W, North K, Kim E, et al. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. *Cancer Cell*. 2018/08/13/ 2018;34(2):225-241.e8. doi:https://doi.org/10.1016/j.ccell.2018.07.003
215. Wadugu BA, Nonavinkere Srivatsan S, Heard A, et al. U2af1 is a haplo-essential gene required for hematopoietic cancer cell survival in mice. *J Clin Invest*. Nov 1 2021;131(21)doi:10.1172/jci141401
216. Madan V, Kanojia D, Li J, et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat Commun*. Jan 14 2015;6:6042. doi:10.1038/ncomms7042
217. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. Sep 11 2011;478(7367):64-9. doi:10.1038/nature10496
218. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011;365(15):1384-1395. doi:10.1056/NEJMoa1103283
219. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts. *New England Journal of Medicine*. 2011/10/13 2011;365(15):1384-1395. doi:10.1056/NEJMoa1103283
220. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. 2022/07/01 2022;36(7):1703-1719. doi:10.1038/s41375-022-01613-1
221. Martin M, Maßhöfer L, Temming P, et al. Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat Genet*. Aug 2013;45(8):933-6. doi:10.1038/ng.2674
222. DeBoever C, Ghia EM, Shepard PJ, et al. Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput Biol*. Mar 2015;11(3):e1004105. doi:10.1371/journal.pcbi.1004105
223. Darman RB, Seiler M, Agrawal AA, et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep*. Nov 3 2015;13(5):1033-45. doi:10.1016/j.celrep.2015.09.053
224. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology*. 2009/02/01 2009;16(2):107-113. doi:10.1038/nsmb.1550

225. Shiozawa Y, Malcovati L, Gallì A, et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nature Communications*. 2018/09/07 2018;9(1):3649. doi:10.1038/s41467-018-06063-x
226. Tang AD, Soulette CM, van Baren MJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nature Communications*. 2020/03/18 2020;11(1):1438. doi:10.1038/s41467-020-15171-6
227. Mian SA, Rouault-Pierre K, Smith AE, et al. SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. *Nature Communications*. 2015/12/08 2015;6(1):10004. doi:10.1038/ncomms10004
228. Mortera-Blanco T, Dimitriou M, Woll PS, et al. SF3B1-initiating mutations in MDS-RSs target lymphomyeloid hematopoietic stem cells. *Blood*. Aug 17 2017;130(7):881-890. doi:10.1182/blood-2017-03-776070
229. Ciurea SO, Merchant D, Mahmud N, et al. Pivotal contributions of megakaryocytes to the biology of idiopathic myelofibrosis. *Blood*. Aug 1 2007;110(3):986-93. doi:10.1182/blood-2006-12-064626
230. Wen QJ, Yang Q, Goldenson B, et al. Targeting megakaryocytic-induced fibrosis in myeloproliferative neoplasms by AURKA inhibition. *Nat Med*. 2015;21(12):1473-1480. doi:10.1038/nm.3995
231. Schischlik F, Jäger R, Rosebrock F, et al. Mutational landscape of the transcriptome offers putative targets for immunotherapy of myeloproliferative neoplasms. *Blood*. 2019;134(2):199-210. doi:10.1182/blood.2019000519
232. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*. Apr 24 2019;8(4):329-337.e4. doi:10.1016/j.cels.2019.03.003
233. Petti AA, Williams SR, Miller CA, et al. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nat Commun*. Aug 14 2019;10(1):3660. doi:10.1038/s41467-019-11591-1
234. Veerappan A, O'Connor NJ, Brazin J, et al. Mast cells: a pivotal role in pulmonary fibrosis. *DNA Cell Biol*. Apr 2013;32(4):206-18. doi:10.1089/dna.2013.2005
235. Legere SA, Haidl ID, Légaré JF, Marshall JS. Mast Cells in Cardiac Fibrosis: New Insights Suggest Opportunities for Intervention. *Front Immunol*. 2019;10:580. doi:10.3389/fimmu.2019.00580
236. Holdsworth SR, Summers SA. Role of Mast Cells in Progressive Renal Diseases. *Journal of the American Society of Nephrology*. 2008;19(12):2254-2261. doi:10.1681/asn.2008010015
237. Huang S, Wu H, Luo F, et al. Exploring the role of mast cells in the progression of liver disease. Review. *Frontiers in Physiology*. 2022-September-13 2022;13doi:10.3389/fphys.2022.964887
238. Piersma B, Hayward MK, Weaver VM. Fibrosis and cancer: A strained relationship. *Biochim Biophys Acta Rev Cancer*. Apr 2020;1873(2):188356. doi:10.1016/j.bbcan.2020.188356
239. Overed-Sayer C, Rapley L, Mustelin T, Clarke D. Are mast cells instrumental for fibrotic diseases? Review. *Frontiers in Pharmacology*. 2014-January-21 2014;4doi:10.3389/fphar.2013.00174
240. Liu B, Yang MQ, Yu TY, et al. Mast Cell Tryptase Promotes Inflammatory Bowel Disease-Induced Intestinal Fibrosis. *Inflamm Bowel Dis*. Jan 19 2021;27(2):242-255. doi:10.1093/ibd/izaa125

241. Kondo S, Kagami S, Kido H, Strutz F, MÜLler GA, Kuroda Y. Role of Mast Cell Tryptase in Renal Interstitial Fibrosis. *Journal of the American Society of Nephrology*. 2001;12(8):1668. doi:10.1681/ASN.V1281668
242. Gunjigake K, Kinoshita J, Yamaguchi T, et al. Interleukin-17A derived from mast cells contributes to fibrosis in gastric cancer with peritoneal dissemination. *Gastric Cancer*. Jan 2021;24(1):31-44. doi:10.1007/s10120-020-01092-2
243. Reiter A, George TI, Gotlib J. New developments in diagnosis, prognostication, and treatment of advanced systemic mastocytosis. *Blood*. 2020;135(16):1365-1376. doi:10.1182/blood.2019000932
244. Doke T, Abedini A, Aldridge DL, et al. Single-cell analysis identifies the interaction of altered renal tubules with basophils orchestrating kidney fibrosis. *Nature Immunology*. 2022/06/01 2022;23(6):947-959. doi:10.1038/s41590-022-01200-7
245. Smyth I, Du X, Taylor MS, Justice MJ, Beutler B, Jackson IJ. The extracellular matrix gene Frem1 is essential for the normal adhesion of the embryonic epidermis. *Proceedings of the National Academy of Sciences*. 2004;101(37):13560-13565. doi:doi:10.1073/pnas.0402760101
246. Ishikawa G, Fujiwara N, Hirschfield H, et al. Shared and Tissue-Specific Expression Signatures between Bone Marrow from Primary Myelofibrosis and Essential Thrombocythemia. *Experimental Hematology*. 2019/11/01/ 2019;79:16-25.e3. doi:https://doi.org/10.1016/j.exphem.2019.10.001
247. Wang M, Guo J, Zhang L, Kuek V, Xu J, Zou J. Molecular structure, expression, and functional role of Clec11a in skeletal biology and cancers. *Journal of Cellular Physiology*. 2020;235(10):6357-6365. doi:https://doi.org/10.1002/jcp.29600
248. Hiraoka A, Sugimura A, Seki T, et al. Cloning, expression, and characterization of a cDNA encoding a novel human growth factor for primitive hematopoietic progenitor cells. *Proceedings of the National Academy of Sciences*. 1997;94(14):7577-7582. doi:doi:10.1073/pnas.94.14.7577
249. Yue R, Shen B, Morrison SJ. Clec11a/osteolectin is an osteogenic growth factor that promotes the maintenance of the adult skeleton. *eLife*. 2016/12/13 2016;5:e18782. doi:10.7554/eLife.18782
250. Shen B, Vardy K, Hughes P, et al. Integrin alpha11 is an Osteolectin receptor and is required for the maintenance of adult skeletal bone mass. *Elife*. Jan 11 2019;8doi:10.7554/eLife.42274
251. Yin C, Zhang J, Guan W, et al. High Expression of CLEC11A Predicts Favorable Prognosis in Acute Myeloid Leukemia. *Front Oncol*. 2021;11:608932. doi:10.3389/fonc.2021.608932
252. Adnan Awad S, Kankainen M, Ojala T, et al. Mutation accumulation in cancer genes relates to nonoptimal outcome in chronic myeloid leukemia. *Blood Advances*. 2020;4(3):546-559. doi:10.1182/bloodadvances.2019000943
253. Giustacchini A, Thongjuea S, Barkas N, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nature Medicine*. 2017/06/01 2017;23(6):692-702. doi:10.1038/nm.4336
254. Kong Y, Zhao S, Tian H, Hai Y. GAS2 Promotes Cell Proliferation and Invasion and Suppresses Apoptosis in Pediatric T-Cell Acute Lymphoblastic Leukemia and Activates Wnt/β-Catenin Pathway. *Onco Targets Ther*. 2020;13:1099-1108. doi:10.2147/ott.S236854

255. Toren A, Bielorai B, Jacob-Hirsch J, et al. CD133-positive hematopoietic stem cell "stemness" genes contain many genes mutated or abnormally expressed in leukemia. *Stem Cells*. Sep 2005;23(8):1142-53. doi:10.1634/stemcells.2004-0317
256. Hernández-Malmierca P, Vonficht D, Schnell A, et al. Antigen presentation safeguards the integrity of the hematopoietic stem cell pool. *Cell Stem Cell*. May 5 2022;29(5):760-775.e10. doi:10.1016/j.stem.2022.04.007
257. Palii CG, Cheng Q, Gillespie MA, et al. Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell*. May 2 2019;24(5):812-820.e5. doi:10.1016/j.stem.2019.02.006
258. Jaiswal S, Jamieson CH, Pang WW, et al. CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell*. Jul 23 2009;138(2):271-85. doi:10.1016/j.cell.2009.05.046
259. Nonino A, Nascimento JM, Mascarenhas CC, Mazzeu JF, Pereira RW, Jacomo RH. CD47 expression is decreased in hematopoietic progenitor cells in patients with myelofibrosis. *Braz J Med Biol Res*. Dec 10 2018;52(1):e7784. doi:10.1590/1414-431x20187784
260. Boasman K, Simmonds MJ, Rinaldi CR. Expression of CD47 and Calr in Myeloproliferative Neoplasms and Myelodysplastic Syndrome: Potential New Therapeutical Targets. *Blood*. 2019;134(Supplement_1):5031-5031. doi:10.1182/blood-2019-128422
261. Chao MP, Takimoto CH, Feng DD, et al. Therapeutic Targeting of the Macrophage Immune Checkpoint CD47 in Myeloid Malignancies. *Front Oncol*. 2019;9:1380. doi:10.3389/fonc.2019.01380
262. Zitvogel L, Tesniere A, Kroemer G. Cancer despite immunosurveillance: immunoselection and immunosubversion. *Nature Reviews Immunology*. 2006/10/01 2006;6(10):715-727. doi:10.1038/nri1936
263. Strickland M, Quek L, Psaila B. The immune landscape in BCR-ABL negative myeloproliferative neoplasms: inflammation, infections and opportunities for immunotherapy. *Br J Haematol*. Mar 2022;196(5):1149-1158. doi:10.1111/bjh.17850
264. Fernandes de Oliveira Costa A, Olops Marani L, Mantello Bianco T, et al. Altered distribution and function of NK-cell subsets lead to impaired tumor surveillance in JAK2V617F myeloproliferative neoplasms. *Front Immunol*. 2022;13:768592. doi:10.3389/fimmu.2022.768592
265. Ivo V, Sanja P, Taghi M, Graciela MN-G, Srdan V, Zeev E. Altered T-cell subset repertoire affects treatment outcome of patients with myelofibrosis. *Haematologica*. 07/30 2020;106(9):2384-2396. doi:10.3324/haematol.2020.249441
266. Leiva O, Ng SK, Chitalia S, Balduini A, Matsuura S, Ravid K. The role of the extracellular matrix in primary myelofibrosis. *Blood Cancer Journal*. 2017/02/01 2017;7(2):e525-e525. doi:10.1038/bcj.2017.6
267. Dunbar A, Kim D, Lu M, et al. CXCL8/CXCR2 signaling mediates bone marrow fibrosis and represents a therapeutic target in myelofibrosis. *bioRxiv*. 2021:2021.12.08.471791. doi:10.1101/2021.12.08.471791
268. Tsai FY, Orkin SH. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood*. May 15 1997;89(10):3636-43.
269. Li Y, Qi X, Liu B, Huang H. The STAT5–GATA2 Pathway Is Critical in Basophil and Mast Cell Differentiation and Maintenance. *The Journal of Immunology*. 2015;194(9):4328-4338. doi:10.4049/jimmunol.1500018

270. Cai M, Langer EM, Gill JG, et al. Dual actions of Meis1 inhibit erythroid progenitor development and sustain general hematopoietic cell proliferation. *Blood*. 2012;120(2):335-346. doi:10.1182/blood-2012-01-403139
271. Nam AS, Kim KT, Chaligne R, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*. Jul 2019;571(7765):355-360. doi:10.1038/s41586-019-1367-0
272. Prchal JF, Axelrad AA. Letter: Bone-marrow responses in polycythemia vera. *N Engl J Med*. Jun 13 1974;290(24):1382. doi:10.1056/nejm197406132902419
273. Lieu YK, Liu Z, Ali AM, et al. SF3B1 mutant-induced missplicing of MAP3K7 causes anemia in myelodysplastic syndromes. *Proc Natl Acad Sci U S A*. Jan 4 2022;119(1)doi:10.1073/pnas.2111703119
274. Mupo A, Seiler M, Sathiaseelan V, et al. Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia*. Mar 2017;31(3):720-727. doi:10.1038/leu.2016.251
275. Lieu YK, Liu Z, Ali AM, et al. SF3B1 mutant-induced missplicing of MAP3K7 causes anemia in myelodysplastic syndromes. *Proceedings of the National Academy of Sciences*. 2022;119(1):e2111703119. doi:doi:10.1073/pnas.2111703119
276. Kanagal-Shamanna R, Adema V, Ma F, et al. SF3B1-Mutant Myelodysplastic Syndrome with Ringed Sideroblasts (MDS-RS) at the Single-Cell Level. *Blood*. 2020/11/05/ 2020;136:50-51. doi:https://doi.org/10.1182/blood-2020-141928
277. Ahmed A, Powers MP, Youker KA, et al. Mast cell burden and reticulin fibrosis in the myeloproliferative neoplasms: A computer-assisted image analysis study. *Pathology - Research and Practice*. 2009/09/15/ 2009;205(9):634-638. doi:https://doi.org/10.1016/j.prp.2009.04.003
278. Keski H. Association of mast cells and bone marrow reticulin fibrosis in patients with bcr-abl negative chronic myeloproliferative neoplasms. *Blood Cells Mol Dis*. May 2020;82:102420. doi:10.1016/j.bcmd.2020.102420
279. Dunphy CH. Evaluation of mast cells in myeloproliferative disorders and myelodysplastic syndromes. *Arch Pathol Lab Med*. Feb 2005;129(2):219-22. doi:10.5858/2005-129-219-eomcim
280. Melo-Cardenas J, Bezavada L, Crawford JC, et al. IL13 Contributes to Fibrotic Progression of the Myeloproliferative Neoplasms. *Blood*. 2021;138(Supplement 1):60-60. doi:10.1182/blood-2021-151492
281. Overed-Sayer C, Miranda E, Dunmore R, et al. Inhibition of mast cells: a novel mechanism by which nintedanib may elicit anti-fibrotic effects. *Thorax*. 2020;75(9):754-763. doi:10.1136/thoraxjnl-2019-214000
282. Chiu A, Nanaji NM, Czader M, et al. The stromal composition of mast cell aggregates in systemic mastocytosis. *Modern Pathology*. 2009/07/01 2009;22(7):857-865. doi:10.1038/modpathol.2009.53
283. Tong J, Sun T, Ma S, et al. Hematopoietic Stem Cell Heterogeneity Is Linked to the Initiation and Therapeutic Response of Myeloproliferative Neoplasms. *Cell Stem Cell*. 2021/03/04/ 2021;28(3):502-513.e6. doi:https://doi.org/10.1016/j.stem.2021.01.018
284. Rao TN, Hansen N, Stetka J, et al. JAK2-V617F and interferon-α induce megakaryocyte-biased stem cells characterized by decreased long-term functionality. *Blood*. Apr 22 2021;137(16):2139-2151. doi:10.1182/blood.2020005563

285. Pellagatti A, Armstrong RN, Steeples V, et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood*. Sep 20 2018;132(12):1225-1240. doi:10.1182/blood-2018-04-843771
286. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia*. 2015/05/01 2015;29(5):1092-1103. doi:10.1038/leu.2014.331
287. Lee SC, North K, Kim E, et al. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. *Cancer Cell*. Aug 13 2018;34(2):225-241.e8. doi:10.1016/j.ccell.2018.07.003
288. Dolatshad H, Pellagatti A, Liberante FG, et al. Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia*. 2016/12/01 2016;30(12):2322-2331. doi:10.1038/leu.2016.149
289. Liu Z, Yoshimi A, Wang J, et al. Mutations in the RNA Splicing Factor SF3B1 Promote Tumorigenesis through MYC Stabilization. *Cancer Discov*. Jun 2020;10(6):806-821. doi:10.1158/2159-8290.Cd-19-1330
290. Liu B, Liu Z, Chen S, et al. Mutant SF3B1 promotes AKT- and NF-κB-driven mammary tumorigenesis. *J Clin Invest*. Jan 4 2021;131(1)doi:10.1172/jci138315
291. Ochi T, Fujiwara T, Ono K, et al. Exploring the mechanistic link between SF3B1 mutation and ring sideroblast formation in myelodysplastic syndrome. *Scientific Reports*. 2022/08/26 2022;12(1):14562. doi:10.1038/s41598-022-18921-2
292. Clough CA, Pangallo J, Sarchi M, et al. Coordinated missplicing of TMEM14C and ABCB7 causes ring sideroblast formation in SF3B1-mutant myelodysplastic syndrome. *Blood*. 2022;139(13):2038-2049. doi:10.1182/blood.2021012652
293. Hershberger CE, Moyer DC, Adema V, et al. Complex landscape of alternative splicing in myeloid neoplasms. *Leukemia*. 2021;35(4):1108-1120. doi:10.1038/s41375-020-1002-y
294. Inoue D, Chew G-L, Liu B, et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature*. 2019/10/01 2019;574(7778):432-436. doi:10.1038/s41586-019-1646-9
295. Lu T, Park S, Zhu J, et al. Overcoming Expressional Drop-outs in Lineage Reconstruction from Single-Cell RNA-Sequencing Data. *Cell Reports*. 2021/01/05/ 2021;34(1):108589. doi:https://doi.org/10.1016/j.celrep.2020.108589
296. Godfrey AL, Chen E, Pagano F, et al. JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. *Blood*. Sep 27 2012;120(13):2704-7. doi:10.1182/blood-2012-05-431791
297. Mylonas E, Yoshida K, Frick M, et al. Single-cell analysis based dissection of clonality in myelofibrosis. *Nature Communications*. 2020/01/07 2020;11(1):73. doi:10.1038/s41467-019-13892-x
298. Brush MH, Weiser DC, Shenolikar S. Growth arrest and DNA damage-inducible protein GADD34 targets protein phosphatase 1 alpha to the endoplasmic reticulum and promotes dephosphorylation of the alpha subunit of eukaryotic translation initiation factor 2. *Mol Cell Biol*. Feb 2003;23(4):1292-303. doi:10.1128/mcb.23.4.1292-1303.2003
299. Monkley S, Overed-Sayer C, Parfrey H, et al. Sensitization of the UPR by loss of PPP1R15A promotes fibrosis and senescence in IPF. *Sci Rep*. Nov 3 2021;11(1):21584. doi:10.1038/s41598-021-00769-7

300. Liu B, Liu Z, Chen S, et al. Mutant SF3B1 promotes AKT- and NF-κB–driven mammary tumorigenesis. *The Journal of Clinical Investigation*. 01/04/ 2021;131(1)doi:10.1172/JCI138315
301. Liu Z, Yoshimi A, Wang J, et al. Mutations in the RNA Splicing Factor SF3B1 Promote Tumorigenesis through MYC Stabilization. *Cancer Discovery*. 2020;10(6):806-821. doi:10.1158/2159-8290.Cd-19-1330
302. Furney SJ, Pedersen M, Gentien D, et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov*. Oct 2013;3(10):1122-1129. doi:10.1158/2159-8290.Cd-13-0330
303. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018/10/01 2018;562(7728):526-531. doi:10.1038/s41586-018-0623-z
304. Jacob AG, Smith CWJ. Intron retention as a component of regulated gene expression programs. *Hum Genet*. Sep 2017;136(9):1043-1057. doi:10.1007/s00439-017-1791-x
305. Baldridge MT, King KY, Boles NC, Weksberg DC, Goodell MA. Quiescent haematopoietic stem cells are activated by IFN-γ in response to chronic infection. *Nature*. 2010/06/01 2010;465(7299):793-797. doi:10.1038/nature09135
306. Walz C, Ahmed W, Lazarides K, et al. Essential role for Stat5a/b in myeloproliferative neoplasms induced by BCR-ABL1 and JAK2V617F in mice. *Blood*. 2012;119(15):3550-3560. doi:10.1182/blood-2011-12-397554
307. Yan D, Jobe F, Hutchison RE, Mohi G. Deletion of Stat3 enhances myeloid cell expansion and increases the severity of myeloproliferative neoplasms in Jak2V617F knock-in mice. *Leukemia*. Oct 2015;29(10):2050-61. doi:10.1038/leu.2015.116
308. Grisouard J, Shimizu T, Duek A, et al. Deletion of Stat3 in hematopoietic cells enhances thrombocytosis and shortens survival in a JAK2-V617F mouse model of MPN. *Blood*. 2015;125(13):2131-2140. doi:10.1182/blood-2014-08-594572
309. Duek A, Lundberg P, Shimizu T, et al. Loss of Stat1 decreases megakaryopoiesis and favors erythropoiesis in a JAK2-V617F-driven mouse model of MPNs. *Blood*. Jun 19 2014;123(25):3943-50. doi:10.1182/blood-2013-07-514208
310. O'Sullivan J, Taylor JA, Gerds AT, et al. Molecular Analysis in the Pacritinib Dose-Finding PAC203 Study in Patients with Myelofibrosis Refractory or Intolerant to Ruxolitinib. *Blood*. 2019;134(Supplement_1):4214-4214. doi:10.1182/blood-2019-129254
311. Jyoti N, Francesca LN, David CW, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica*. 11/01 2015;100(11):e438-e442. doi:10.3324/haematol.2015.129510
312. Ortmann CA, Kent DG, Nangalia J, et al. Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med*. Feb 12 2015;372(7):601-612. doi:10.1056/NEJMoa1412098
313. Al Assaf C, Van Obbergh F, Billiet J, et al. Analysis of phenotype and outcome in essential thrombocythemia with CALR or JAK2 mutations. *Haematologica*. Jul 2015;100(7):893-7. doi:10.3324/haematol.2014.118299
314. Rodriguez-Meira A, Norfo R, Wen WX, et al. Deciphering TP53 mutant Cancer Evolution with Single-Cell Multi-Omics. *bioRxiv*. 2022:2022.03.28.485984. doi:10.1101/2022.03.28.485984
315. Picelli S, Björklund Å K, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. Nov 2013;10(11):1096-8. doi:10.1038/nmeth.2639

316. Hagemann-Jensen M, Ziegenhain C, Chen P, et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology*. 2020/06/01 2020;38(6):708-714. doi:10.1038/s41587-020-0497-0
317. Melo-Cardenas J, Bezavada L, Crawford JC, et al. IL-13/IL-4 signaling contributes to fibrotic progression of the myeloproliferative neoplasms. *Blood*. Oct 25 2022;doi:10.1182/blood.2022017326
318. Nakayama S, Yokote T, Hiraoka N, et al. Transforming growth factor β- and interleukin 13-producing mast cells are associated with fibrosis in bone marrow. *Hum Pathol*. Apr 2017;62:180-186. doi:10.1016/j.humpath.2017.01.007
319. Leidner R, Sanjuan Silva N, Huang H, et al. Neoantigen T-Cell Receptor Gene Therapy in Pancreatic Cancer. *New England Journal of Medicine*. 2022;386(22):2112-2119. doi:10.1056/NEJMoa2119662
320. Zhao X, Pan X, Wang Y, Zhang Y. Targeting neoantigens for cancer immunotherapy. *Biomarker Research*. 2021/07/28 2021;9(1):61. doi:10.1186/s40364-021-00315-7
321. Biernacki MA, Foster KA, Clough C, et al. A Shared SF3B1 Neoantigen Is Presented on Primary Malignant Cells and Induced Pluripotent Stem Cell-Derived Hematopoietic Lines. *Blood*. 2020;136(Supplement 1):13-14. doi:10.1182/blood-2020-139907

Appendix 1

Primer sequences for ADT/HTO PCR amplification

Primer name	Sequence
SI-PCR primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC*T*C
HTO additive primer v2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
ADT additive primer	CCTTGGCACCCGAGAATT*C*C
D701_LONG	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
D702_LONG	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
D703_LONG	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
D704_LONG	CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
RPI1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTC*C*A
RPI2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTC*C*A
RPI3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTC*C*A
RPI4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTC*C*A

Appendix 2

Primer sequences for TARGET-seq RT-PCR step

Step	Region (hg19 coordinates)	Gene	Forward Primer	Reverse Primer
RT	chr9:5069183-5077565	*JAK2*	TAAATGCTGTCCCCCAAAGC	CCATGCCAACTGTTTAGCAAC
RT	chr2:198266772-198267754	*SF3B1*	ACCATAAGGAGTTGCTGCTTCA	CGTGGTCATTGAACCGCTATTG
PCR (cDNA)	chr9:5069183-5077565	*JAK2*	AAGCAGTGGTATCAACGCAGAGTTAAATGCTGTCCCCCAAAGC	AAGCAGTGGTATCAACGCAGAGTCCATGCCAACTGTTTAGCAAC
PCR (cDNA)	chr2:198266772-198267754	*SF3B1*	AAGCAGTGGTATCAACGCAGAGTACCATAAGGAGTTGCTGCTTCA	AAGCAGTGGTATCAACGCAGAGTCGTGGTCATTGAACCGCTATTG
PCR (gDNA)	chr9:5073560-5073881	*JAK2*	CCAAAGCACATTGTATCCTCATCT	CACTGACACCTAGCTGTGATCCT
PCR (gDNA)	chr2:198267075-198267591	*SF3B1*	ATCTCCCCAAATCAGTAGCCCAA	ATTACCAACTCATGACTGTCCTTTC

Appendix 3

Targeted forward primer sequences for TARGET-seq PCR1 step (with plate barcodes)

Step	Region (hg19 coordinates)	Gene	Barcode name	CS1	Barcode	Target sequence
PCR1 (cDNA)	chr9:5072596-5077484	*JAK2*	L07	ACACTGACGACATGGTTCTACA	CCTGGA	TCTGGATAAAGCACACAGAAACT
PCR1 (cDNA)	chr9:5072596-5077484	*JAK2*	L08	ACACTGACGACATGGTTCTACA	TGACTA	TCTGGATAAAGCACACAGAAACT
PCR1 (cDNA)	chr2:198266818 -198267713	*SF3B1*	L07	ACACTGACGACATGGTTCTACA	CCTGGA	CTGATGGTCCGAACTTTCTGCT
PCR1 (cDNA)	chr2:198266818 -198267713	*SF3B1*	L08	ACACTGACGACATGGTTCTACA	TGACTA	CTGATGGTCCGAACTTTCTGCT

PCR1 (gDNA)	chr9:5073648-5073817	*JAK2*	L07	ACACTGACGACATG GTTCTACA	CCTGGA	TTATGGACAACAGTCAAA CAACAA
PCR1 (gDNA)	chr9:5073648-5073817	*JAK2*	L08	ACACTGACGACATG GTTCTACA	TGACTA	TTATGGACAACAGTCAAA CAACAA
PCR1 (gDNA)	chr2:198267118 -198267465	*SF3B1*	L07	ACACTGACGACATG GTTCTACA	CCTGGA	TTAAACATGGACAGGCT GTGTG
PCR1 (gDNA)	chr2:198267118 -198267465	*SF3B1*	L08	ACACTGACGACATG GTTCTACA	TGACTA	TTAAACATGGACAGGCT GTGTG

Targeted reverse primer sequences for TARGET-seq PCR1 step (with plate barcodes)

Step	Region (hg19 coordinates)	Gene	Barcode name	CS2	Barcode	Target sequence
PCR1 (cDNA)	chr9:5072596-5077484	*JAK2*	R07	TACGGTAGCAGAGA CTTGGTCT	CGTGTT	TCCAAATTTTACAAACTC CTGAACC
PCR1 (cDNA)	chr9:5072596-5077484	*JAK2*	R08	TACGGTAGCAGAGA CTTGGTCT	CGCTCA	TCCAAATTTTACAAACTC CTGAACC
PCR1 (cDNA)	chr2:198266818 - 198267713	*SF3B1*	R07	TACGGTAGCAGAGA CTTGGTCT	CGTGTT	AGAGTGGAAGGCCGAGA GAT
PCR1 (cDNA)	chr2:198266818 - 198267713	*SF3B1*	R08	TACGGTAGCAGAGA CTTGGTCT	CGCTCA	AGAGTGGAAGGCCGAGA GAT
PCR1 (gDNA)	chr9:5073648-5073817	*JAK2*	R07	TACGGTAGCAGAGA CTTGGTCT	CGTGTT	AAAGGCATTAGAAAGCCT GTAGT
PCR1 (gDNA)	chr9:5073648-5073817	*JAK2*	R08	TACGGTAGCAGAGA CTTGGTCT	CGCTCA	AAAGGCATTAGAAAGCCT GTAGT
PCR1 (gDNA)	chr2:198267118 - 198267465	*SF3B1*	R07	TACGGTAGCAGAGA CTTGGTCT	CGTGTT	CTTTTGCTGTTGTAGCCT CTGC
PCR1 (gDNA)	chr2:198267118 - 198267465	*SF3B1*	R08	TACGGTAGCAGAGA CTTGGTCT	CGCTCA	CTTTTGCTGTTGTAGCCT CTGC

Appendix 4

TARGET-seq sequencing primers (plus sign denotes LNA-modified bases); all primers were HPLC-purified

Primer name	Sequence
TSO-LNA (RNase-free)	AAGCAGTGGTATCAACGC AGAGTACATrGrG+G
ISPC	AAGCAGTGGTATCAACGCAGAGT
CS1 sequencing primer	A+CA+CTG+ACGACATGGTTCTACA
CS2 sequencing primer	T+AC+GGT+AGCAGAGACTTGGTCT
LCS1 sequencing primer	GGCGACCACCGAGATCTAC ACTGACGACATGGTTCTACA
CS1rc sequencing primer	T+GT+AG+AACCATGTCGTCAGTGT
CS2rc sequencing primer	A+GAC+CA+AGTCTCTGCTACCGTA

Appendix 5

Baseline mutations in the MAJICET cohort

ID	Gene	Genomic position	Nucleotide change	Protein consequence	Mutation type	VAF (%)
35	*TET2*	4:106157960	c.2861G>A	p.Trp954Xaa	SNV; nonsense	52.39
35	*TET2*	4:106164860	c.3729_3733delACTCT	p.Tyr1245GlyfsTer21	Frameshift del	6.18
35	*TET2*	4:106193999	c.4462delA	p.Asn1489MetfsXaa82	Frameshift del	17.74
54	*TET2*	4:106155920	c.822delC	p.Asn275IlefsXaa18	Frameshift del	38.55

55	*TET2*	4:106197236	c.5570_5571delCT	p.Pro1857ArgfsXaa17	Frameshift del	7.06
106	*TET2*	4:106155301	c.202G>T	p.Gly68Xaa	SNV; nonsense	37
108	*TET2*	4:106164939	c.3803+4A>T		Frameshift ins	28.86
128	*TET2*	4:106156540	c.1441C>T	p.Gln481Xaa	SNV; nonsense	41.35
131	*TET2*	4:106180796	c.3824G>T	p.Gly1275Val	SNV; missense	33.92
148	*TET2*	4:106162586	c.3500G>A	p.Arg1167Lys	SNV; missense	28.4
180	*TET2*	4:106196446	c.4781del	p.Pro1594LeufsXaa2	Frameshift del	16.61
184	*TET2*	4:106156365	c.1270del	p.Ser424AlafsXaa3	Frameshift del	5.47
213	*TET2*	4:106180829	c.3857C>T	p.Ser1286Phe	SNV; missense	34.35
213	*TET2*	4:106180854	c.3882C>G	p.Tyr1294TXaa	SNV; nonsense	40.71
215	*TET2*	4:106156812	c.1716dup	p.His573SerfsTer10	Frameshift ins	17.79
35	*SF3B1*	2:198266834	c.2098A>G	p.Lys700Glu	SNV; missense	46.58
73	*SF3B1*	2:198267483	c.1874G>T	p.Arg625Leu	SNV; missense	28.77
87	*SF3B1*	2:198267359	c.1998G>C	p.Lys666Asn	SNV; missense	43.62
103	*SF3B1*	2:198267359	c.1998G>C	p.Lys666Asn	SNV; missense	41.33
151	*SF3B1*	2:198267359	c.1998G>C	p.Lys666Asn	SNV; missense	40.38
193	*SF3B1*	2:198267359	c.1998G>C	p.Lys666Asn	SNV; missense	19.56
123	*SF3B1*	2:198267360	c.1997A>G	p.Lys666Arg	SNV; missense	17.2
27	*TP53*	17:7578406	c.524G>A	p.Arg175His	SNV; missense	19.8
38	*TP53*	17:7577545	c.736A>G	p.Met246Val	SNV; missense	38.46
118	*TP53*	17:7577578	c.703A>G	p.Asn235Asp	SNV; missense	5.09
118	*TP53*	17:7577568	c.713G>A	p.Cys238Tyr	SNV; missense	13.01
131	*TP53*	17:7577556	c.725G>A	p.Cys242Tyr	SNV; missense	16.98
158	*TP53*	17:7577157	c.783-2A>C		Splice site	8.79
180	*TP53*	17:7578495	c.433_435del	p.Leu145del	Frameshift del	19.3
184	*TP53*	17:7577120	c.818G>A	p.Arg273His	SNV; missense	11.06
25	*ASXL1*	20:31022442	c.1934dup	p.Gly646TrpfsXaa12	Frameshift ins	28.64
115	*ASXL1*	20:31022442	c.1934dup	p.Gly646TrpfsXaa12	Frameshift ins	22.7
125	*ASXL1*	20:31022442	c.1934dup	p.Gly646TrpfsXaa12	Frameshift ins	27.72
130	*ASXL1*	20:31023408	c.2893C>T	p.Arg965Xaa	SNV; missense	9.03
196	*ASXL1*	20:31022793	c.2278C>T	p.Gln760Xaa	SNV; nonsense	38.79
61	*DNMT3A*	2:25457242	c.2645G>A	p.Arg882His	SNV; missense	15.69
128	*DNMT3A*	2:25457242	c.2645G>A	p.Arg882His	SNV; missense	37.12
198	*DNMT3A*	2:25457278	c.2603_2609del	p.Phe868SerfsXaa11	Frameshift del	23.97
57	*IDH2*	15:90631934	c.419G>A	p.Arg140Gln	SNV; missense	46.03
193	*IDH2*	15:90631934	c.419G>A	p.Arg140Gln	SNV; missense	19.22

101	*ZRSR2*	X:15821921	c.312+2T>A		Frameshift ins	10.98
104	*ZRSR2*	X:15838370	c.868C>T	p.Arg290Xaa	SNV; nonsense	92.73
61	*ETV6*	12:12006434	c.403del	p.His135ThrfsXaa74	Frameshift del	16.02
196	*PHF6*	X:133511706	c.59_60insT	p.Lys21Xaa	Frameshift ins	74.96
106	*SRSF2*	17:74732959	c.286G>A	p.Pro95Leu	SNV; missense	35.29
213	*EZH2*	7:148514322	c.1402T>G	p.Cys468Gly	SNV; missense	20.77
99	*CSF3R*	1:36932076	c.2474G>A	p.Gly825Glu	SNV; missense	39.63

Longitudinal mutations MAJICET cohort

ID	Gene	Genomic position	Nucleotide change	Protein consequence	Mutation type	VAF (%)
55	*TET2*	4:106158481	c.3382dup	p.Tyr1128LeufsXaa2	Frameshift ins	16.68
55	*TET2*	4:106162529	c.3443A>G	p.Tyr1148Cys	SNV; missense	17.29
56	*TET2*	4:106157914	c.2815C>T	p.Gln939Xaa	SNV; nonsense	7.63
118	*TET2*	4:106190831	c.4109G>A	p.Gly1370Glu	SNV; missense	6.15
180	*TET2*	4:106155238	c.141_150del	p.Val48ThrfsXaa16	Frameshift del	6.79
203	*TET2*	4:106164080	c.3590A>G	p.Lys1197Arg	SNV; missense	5.97
55	*ASXL1*	20:31023440	c.2925T>A	p.Cys975Xaa	SNV; nonsense	13.61
55	*ASXL1*	20:31022442	c.1934dup	p.Gly646TrpfsXaa12	Frameshift ins	17.73
67	*ASXL1*	20:31022442	c.1934dup	p.Gly646TrpfsXaa12	Frameshift ins	9.75
76	*ASXL1*	20:31022903	c.2388G>A	p.Trp796Xaa	SNV; nonsense	23.25
89	*TP53*	17:7578206	c.643A>G	p.Ser215Gly	SNV; missense	14.5
128	*TP53*	17:7577568	c.713G>A	p.Cys238Tyr	SNV; missense	6.67
128	*TP53*	17:7579515	c.169_172del	p.Asp57GlnfsXaa65	Frameshift del	7.43
81	*SF3B1*	2:198267359	c.1998G>C	p.Lys666Asn	SNV; missense	5.77
118	*SF3B1*	2:198267359	c.1998G>C	p.Lys666Asn	SNV; missense	5.44
114	*SETBP1*	18:42531907	c.2602G>A	p.Asp868Asn	SNV; missense	36.44
128	*U2AF1*	21:44524456	c.101C>T	p.Ser34Phe	SNV; missense	8.5

Baseline mutations PAC203 cohort

Sample	Variant_Classification	Hugo_Symbol	Chromosome	HGVSp
44001-3092	Nonsense_Mutation	*TET2*	4	NP_001120680.1:p.Gln255Ter
44001-3092	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3089	Nonsense_Mutation	*TET2*	4	NP_001120680.1:p.Gln373Ter
44001-3089	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3089	Missense_mutation	*RUNX1*	21	NP_001745.2:p.Arg162Gly
44001-3089	In_Frame_Del	*SRSF2*	17	NP_001182356.1:p.Pro95_Arg102del
36003-3066	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe

36003-3126	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Thr
36003-3126	Missense_mutation	*EZH2*	7	NP_004447.2:p.Tyr741Cys
36003-3126	Missense_mutation	*EZH2*	7	NP_004447.2:p.Cys554Arg
36003-3126	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
36003-3126	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Glu790Ter
36003-3126	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Lys203ArgfsTer4
36003-3126	Nonsense_Mutation	*EZH2*	7	NP_004447.2:p.Ser644Ter
36003-3100	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
36003-3100	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
36003-3080	Missense_mutation	*EZH2*	7	NP_004447.2:p.Ala682Glu
36003-3080	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
36003-3108	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
36003-3108	Nonsense_Mutation	*TET2*	4	NP_001120680.1:p.Arg1404Ter
36003-3108	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Glu635ArgfsTer15
36003-3082	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Asp
36003-3082	Splice_site	*EZH2*	7	
36003-3082	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
36003-3079	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
36003-3079	Frame_Shift_Del	*ZRSR2*	X	NP_005080.1:p.Lys26AsnfsTer12
36003-3079	Splice_site	*CBL*	11	
34001-3104	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
34001-3104	Missense_mutation	*IDH2*	15	NP_002159.2:p.Arg140Gln
34001-3104	Missense_mutation	*SRSF2*	17	NP_001182356.1:p.Pro95Leu
34001-3104	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Ser
34001-3129	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44007-3107	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44006-3132	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Glu263ArgfsTer30
44006-3132	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44005-3143	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Asn
44005-3143	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10060-3084	Frame_Shift_Ins	*TET2*	4	NP_001120680.1:p.Arg1440ThrfsTer38
10060-3084	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10035-3035	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10060-3102	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10060-3102	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Ser34Phe
10060-3098	Missense_mutation	*MPL*	1	NP_005364.1:p.Ser505Asn
10060-3098	In_Frame_Del	*TP53*	17	NP_000537.3:p.His179del
10060-3098	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10060-3099	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Asp

10060-3099	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10048-3116	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10048-3116	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gln61Arg
10048-3116	Missense_mutation	*SRSF2*	17	NP_001182356.1:p.Pro95His
10048-3116	Missense_mutation	*SETBP1*	18	NP_056374.2:p.Gly870Ser
10015-3078	Missense_mutation	*IDH1*	2	NP_005887.2:p.Arg132His
10015-3078	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10015-3078	In_Frame_Del	*SRSF2*	17	NP_001182356.1:p.Pro95_Arg102del
10015-3078	Missense_mutation	*IDH1*	2	NP_005887.2:p.Arg132His
10015-3078	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10015-3078	In_Frame_Del	*SRSF2*	17	NP_001182356.1:p.Pro95_Arg102del
10025-3052	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10025-3052	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10035-3070	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10035-3070	Missense_mutation	*CBL*	11	NP_005179.2:p.Leu380Pro
10035-3070	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Gln733Ter
10035-3063	Missense_mutation	*TET2*	4	NP_001120680.1:p.Ser1284Ala
10035-3063	Missense_mutation	*TET2*	4	NP_001120680.1:p.Phe1285Ser
10035-3063	Missense_mutation	*EZH2*	7	NP_004447.2:p.Ala682Val
10035-3063	Frame_Shift_Del	*EZH2*	7	NP_004447.2:p.Ile459LeufsTer4
10035-3063	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10035-3063	Missense_mutation	*CBL*	11	NP_005179.2:p.Cys396Tyr
10035-3063	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Ala640GlyfsTer14
10035-3063	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10035-3063	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Leu731TyrfsTer13
10035-3063	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Leu775Ter
10035-3063	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10065-3059	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Asn595LeufsTer41
10065-3059	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10066-3110	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Leu
10062-3061	Frame_Shift_Del	*EZH2*	7	NP_004447.2:p.Pro257LeufsTer8
10062-3061	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3061	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Asp
10062-3076	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Asp
10062-3076	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3076	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Tyr591Ter
10062-3076	Missense_mutation	*PHF6*	X	NP_115834.1:p.Ile314Thr
10062-3056	Missense_mutation	*EZH2*	7	NP_004447.2:p.Phe729Ser
10062-3056	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe

10035-3142	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Leu
10035-3109	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10058-3009	Missense_mutation	*JAK2*	9	NP_004963.1:p.Tyr613His
10058-3009	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3073	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3073	Missense_mutation	*IDH2*	15	NP_002159.2:p.Arg140Gln
10013-3064	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Thr
10013-3064	Missense_mutation	*TET2*	4	NP_001120680.1:p.His1416Asp
10013-3064	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10013-3064	Missense_mutation	*PHF6*	X	NP_115834.1:p.Ile314Thr
10046-3057	Missense_mutation	*SETBP1*	18	NP_056374.2:p.Asp868Asn
10046-3057	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10046-3057	Missense_mutation	*ETV6*	12	NP_001978.1:p.Ser22Ile
10046-3071	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10046-3071	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gln61Pro
10046-3071	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Ser34Phe
10046-3071	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10046-3011	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10046-3011	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Gln1350Ter
10046-3011	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Ser34Phe
10062-3134	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Asn598LysfsTer2
10062-3134	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3134	Missense_mutation	*ZRSR2*	X	NP_005080.1:p.Arg452Cys
10062-3123	Frame_Shift_Ins	*TET2*	4	NP_001120680.1:p.Ser1392GlnfsTer9
10062-3123	Frame_Shift_Ins	*CALR*	19	NP_004334.1:p.Lys385AsnfsTer47
10059-3124	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Glu1250ArgfsTer3
10059-3124	Missense_mutation	*TET2*	4	NP_001120680.1:p.Ile1873Thr
10059-3124	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10059-3124	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Ser34Phe
10059-3145	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10059-3145	Missense_mutation	*TP53*	17	NP_000537.3:p.Cys176Phe
34003-3077	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
34003-3077	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Ala735LeufsTer9
34003-3077	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Ser34Tyr
34003-3077	Missense_mutation	*IDH2*	15	NP_002159.2:p.Arg140Gln
34001-3060	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
39001-3087	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
39001-3174	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
39001-3174	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Gln829Ter

39001-3067	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
39001-3067	Missense_mutation	*TET2*	4	NP_001120680.1:p.Met66Leu
44007-3200	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
44007-3200	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
44007-3192	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Lys
44007-3192	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
44001-3189	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3189	In_Frame_Del	*SRSF2*	17	NP_001182356.1:p.Pro95_Arg102del
44001-3189	Missense_mutation	*SETBP1*	18	NP_056374.2:p.Gly870Ser
33010-3149	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
33010-3181	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Leu
33010-3181	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Val
33010-3181	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
33010-3181	Frame_Shift_Del	*ASXL1*	20	
44005-3183	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
44009-3178	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44009-3178	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Thr717Ala
44009-3178	Missense_mutation	*TET2*	4	NP_001120680.1:p.Glu1207Gly
44009-3178	Missense_mutation	*TP53*	17	NP_000537.3:p.Arg175His
44009-3198	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Asn
44009-3198	In_Frame_Del	*PHF6*	X	NP_115834.1:p.Glu338del
44009-3198	Frame_Shift_Ins	*CALR*	19	NP_004334.1:p.Lys385AsnfsTer47
33001-3177	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
33001-3179	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
39004-3088	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Leu
39004-3088	Missense_mutation	*CBL*	11	NP_005179.2:p.Cys416Ser
39004-3088	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gln61Arg
39004-3088	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
39004-3088	Missense_mutation	*RUNX1*	21	NP_001745.2:p.Ala142Pro
39004-3088	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
39004-3088	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Asp
39004-3088	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gln916Ter
39004-3088	Missense_mutation	*KRAS*	12	NP_203524.1:p.Ala18Asp
33001-3180	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
33001-3180	Missense_mutation	*IDH2*	15	NP_002159.2:p.Arg140Gln
33001-3180	Missense_mutation	*SRSF2*	17	NP_001182356.1:p.Pro95Leu
39001-3196	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Arg
39001-3196	Frame_Shift_Ins	*TET2*	4	NP_001120680.1:p.Thr221AsnfsTer4
39001-3196	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe

39001-3196	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
39001-3190	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Leu
39001-3190	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Leu1213IlefsTer3
39001-3190	Missense_mutation	*EZH2*	7	NP_004447.2:p.Gly660Glu
10068-3184	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Ala744Pro
10068-3184	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10068-3184	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Ser890Ter
10062-3091	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3091	Missense_mutation	*TP53*	17	NP_000537.3:p.Cys176Phe
10035-3201	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gln1547Ter
10035-3201	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10014-3203	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Arg625Leu
10014-3203	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10014-3203	Missense_mutation	*PHF6*	X	NP_115834.1:p.Met278Lys
10039-3187	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Pro419LeufsTer8
10039-3187	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Glu635ArgfsTer15
10039-3187	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
10065-3147	Missense_mutation	*TET2*	4	NP_001120680.1:p.Cys1135Tyr
10065-3147	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10065-3147	Frame_Shift_Del	*PHF6*	X	NP_115834.1:p.Gln121ArgfsTer16
10065-3147	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly13Arg
10046-3199	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Asn
10046-3199	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10046-3159	Missense_mutation	*SF3B1*	2	NP_036565.2:p.His662Asp
10046-3159	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10046-3159	Missense_mutation	*IDH2*	15	NP_002159.2:p.Arg140Gln
10046-3156	Frame_Shift_Ins	*TET2*	4	NP_001120680.1:p.Leu200HisfsTer5
10046-3156	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10049-3128	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10049-3128	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Ser34Phe
10049-3128	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Leu721CysfsTer4
10002-3165	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gln1274Ter
10002-3165	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10002-3165	Splice_site	*TET2*	4	
10002-3165	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Asn1346IlefsTer17
10048-3188	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10048-3188	Missense_mutation	*IDH2*	15	NP_002159.2:p.Arg140Gln
10048-3188	Missense_mutation	*TP53*	17	NP_000537.3:p.Gly266Glu
10048-3188	Missense_mutation	*TP53*	17	NP_000537.3:p.Tyr163Asn

10048-3188	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10048-3188	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10048-3090	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10048-3090	Missense_mutation	*TET2*	4	NP_001120680.1:p.Thr1884Ala
10046-3160	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys700Glu
10046-3160	Frame_Shift_Ins	*CALR*	19	NP_004334.1:p.Lys385AsnfsTer47
10046-3160	Missense_mutation	*TP53*	17	NP_000537.3:p.Arg248Trp
10060-3168	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gln1652Ter
10060-3168	Frame_Shift_Ins	*EZH2*	7	NP_004447.2:p.Asn263GlnfsTer8
10060-3168	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10060-3168	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
10060-3168	Missense_mutation	*IDH1*	2	NP_005887.2:p.Arg132Cys
10060-3168	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gln1652Ter
10060-3168	Frame_Shift_Ins	*EZH2*	7	NP_004447.2:p.Asn263GlnfsTer8
10060-3168	Missense_mutation	*CBL*	11	NP_005179.2:p.Cys396Ser
10060-3168	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10060-3168	Missense_mutation	*CBL*	11	NP_005179.2:p.Cys419Tyr
10060-3168	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
10064-3193	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10064-3193	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10064-3193	Frame_Shift_Ins	*PHF6*	X	p.Ter366Serext*21
10059-3157	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10059-3157	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Asp
10059-3157	Missense_mutation	*RUNX1*	21	NP_001745.2:p.Asp160Tyr
10059-3166	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Arg
10059-3166	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10073-3194	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Thr
10073-3194	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Gly949ValfsTer2
10073-3194	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
10012-3171	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10022-3234	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10022-3234	Missense_mutation	*ATRX*	X	NP_000480.2:p.Arg905Thr
10035-3206	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Leu1816Ter
10035-3206	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10022-3221	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gly12Ser
10022-3221	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gly12Arg
10022-3221	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
10039-3244	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10039-3244	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12

10039-3244	Missense_mutation	*U2AF1*	21	NP_001020374.1:p.Gln157Pro
10046-3213	Missense_mutation	*EZH2*	7	NP_004447.2:p.His158Pro
10046-3213	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10046-3213	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10062-3231	Missense_mutation	*DNMT3A*	2	NP_072046.2:p.Arg882Ser
10062-3231	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys700Glu
10062-3231	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10062-3231	Missense_mutation	*TET2*	4	NP_001120680.1:p.His269Arg
10062-3231	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Gln557SerfsTer4
10062-3231	Frame_Shift_Ins	*TET2*	4	NP_001120680.1:p.Cys1358LeufsTer43
10062-3225	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Lys
10066-3237	Missense_mutation	*MPL*	1	NP_005364.1:p.Trp515Leu
10062-3210	Missense_mutation	*EZH2*	7	NP_004447.2:p.Arg690His
10062-3210	Missense_mutation	*EZH2*	7	NP_004447.2:p.Asn130Lys
10062-3210	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gln61Pro
10062-3210	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Glu635ArgfsTer15
10025-3218	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10025-3218	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10002-3209	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10014-3207	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10014-3207	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
10014-3207	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Ser846GlnfsTer5
10014-3207	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gly12Arg
10014-3207	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Phe1373CysfsTer6
10035-3211	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gln61Arg
10035-3211	Missense_mutation	*ASXL1*	20	NP_056153.2:p.Lys1446Glu
10035-3211	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Glu364ArgfsTer49
44001-3224	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys666Asn
44001-3224	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3228	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Leu1231TrpfsTer22
44001-3228	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3228	Frame_Shift_Del	*TET2*	4	NP_001120680.1:p.Leu1231TrpfsTer22
44001-3228	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3228	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Arg1415Ter
39001-3222	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44009-3215	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
34005-3235	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44001-3227	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gln1537Ter
44001-3227	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Cys1875Ter

44001-3227	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
34001-3230	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
34001-3230	Frame_Shift_Ins	*ASXL1*	20	NP_056153.2:p.Gly646TrpfsTer12
34001-3230	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Arg965Ter
44006-3236	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
44006-3236	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gln61Pro
10063-3243	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
33002-3122	Missense_mutation	*SF3B1*	2	NP_036565.2:p.Lys700Glu
33002-3122	Missense_mutation	*TP53*	17	NP_000537.3:p.Tyr163Cys
33002-3122	Frame_Shift_Del	*CALR*	19	NP_004334.1:p.Leu367ThrfsTer46
46001-3120	Missense_mutation	*NRAS*	1	NP_002515.1:p.Gly12Cys
46001-3120	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10022-3175	Nonsense_mutation	*TET2*	4	NP_001120680.1:p.Gly1735Ter
10022-3175	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10015-3117	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10015-3117	Missense_mutation	*KRAS*	12	NP_203524.1:p.Gly12Val
10015-3117	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Trp1037Ter
44014-3208	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10073-3212	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10073-3212	Frame_Shift_Del	*ASXL1*	20	NP_056153.2:p.Glu635ArgfsTer15
10073-3212	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Arg693Ter
10006-3219	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10006-3219	Missense_mutation	*ETV6*	12	NP_001978.1:p.Asp206Glu
10006-3219	Missense_mutation	*ATRX*	X	NP_000480.2:p.Ser1192Ile
10006-3219	Missense_mutation	*ATRX*	X	NP_000480.2:p.Gln872Arg
10006-3219	Nonsense_mutation	*ATRX*	X	NP_000480.2:p.Gly628Ter
10006-3216	Missense_mutation	*JAK2*	9	NP_004963.1:p.Val617Phe
10006-3216	Missense_mutation	*KRAS*	12	NP_203524.1:p.Tyr64Asp
10006-3216	Nonsense_mutation	*ASXL1*	20	NP_056153.2:p.Gln858Ter
10006-3216	Splice_site	*ZRSR2*	X	