Investigating the importance of monocyte crosstalk with haematopoietic stem and progenitor cells in TET2 mutant chronic myelomonocytic leukaemia



Dr Richard Salisbury

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity term 2025

Acknowledgements

This thesis is an embodiment of my ambition to become a scientist and the fruit of almost 6 years of endeavour. I dedicate this work to those who have helped me to achieve this goal as it would not have been possible without the unerring support of my family and my mentors. In particular I would like to thank my wife, Emily, who has undoubtedly sacrificed the most to allow me to create this work. She has always supported me to follow my dreams in spite of the lonely evenings and weekends when she has lost me to my work. I wish to acknowledge this for what it represents, a true act of love and altruism. I hope there will come a time where I can offer the same steadfast support to you as you have shown to me so I can demonstrate how grateful I am and have been to you.

I would also like to acknowledge the role of my parents in encouraging my academic development, and footing the bill when I decided to go and play around with test tubes during a Masters of Research instead of putting my head down and finishing medical school. The freedom to pursue my curiosity in the natural world has directly led to my ambition and success in pursuing a career in research. I would like to thank my funders, Cancer Research UK, SynOx Therapeutics and the British Society of Haematology for their financial and intellectual support of my work. The Oxford Cancer team, Kea, Emily, Michael and Mark who have created the most wonderful environment in which to pursue a DPhil. The workshops and seminar series provided as part of the Cancer Science programme really helped to frame my work within the bigger picture of cancer research. I would also like to thank the past and present members of the Haematopoietic Stem Cell Laboratory for their support and for making the lab a joyful place to visit. In particular I would like to thank Christina whose guidance and practical assistance was fundamental for the success of many of the critical experiments presented in this thesis.

Finally, I would like to whole heartedly thank my supervisors and mentors. Initially Lynn, who was instrumental to the development of the plan for this research and more latterly Skirmantas, who I hope to continue working with in the years to come. Most of all I would like to thank Adam for taking a chance on me, and who has always been generous with his time and encouragement. His specific advice and guidance has been instrumental to the success of this project but even more than this, experiencing his seemingly insurmountable capacity for work and constant enthusiasm for research has been a source of great inspiration for me. I hope I have been able to repay your trust and belief in me with the work presented in this thesis.

Abstract

0.1 Lay abstract

Chronic myelomonocytic leukaemia (CMML) is a cancer of the blood with no disease specific treatments and is incurable in most people. Scientists do not really understand why people develop CMML. Faulty genes in CMML cells, caused by mistakes in the genetic instruction manual for each cell, are neither specific to CMML nor do they correlate well with how long patients survive. For instance, faulty TET2 genes, leading to a reduced ability to make the TET2 protein, are the most common alteration found in CMML patients.

I studied the role of faulty TET2 genes in CMML by testing how gene expression and the way in which the DNA is folded became altered in CMML blood cells compared to healthy donor cells. I found that haematopoietic stem cells (HSCs), the cells which give rise to all cells in the blood, become primed to produce myeloid immune cells in a similar way to that seen after someone has had a severe infection. In the context of infection, this immune memory is protective and allows the rapid production of immune cells in response to a second infection. However, in CMML it causes the inappropriate creation of immune cells in the absence of infection, using up energy and reducing the body's ability to fight infection when it does occur.

Blood cells are normally created by HSCs that live within the bone marrow. I mapped out changes to the bone marrow of CMML patients and healthy donors to identify potential molecular signals that could trigger myeloid immune cell production by HSCs. CMML patient bone marrow becomes flooded with monocytes and macrophages, two types of myeloid immune blood cell. They produce signals such as M-CSF (macrophage-colony stimulating factor) that can feedback to the HSCs, promoting further myeloid immune blood cell production.

I created a model of these bone marrow interactions by growing HSCs in a dish. I found that blocking the function of M-CSF, using an antibody called emactuzumab, was able to reduce myeloid immune cell production in over half of the CMML patients. I am working in collaboration with the company that makes emactuzumab to check these results further, but we are hopeful that emactuzumab may emerge as an exciting new treatment for patients with CMML.

0.2 Scientific abstract

Background: TET2 loss of function mutations (TET2m) are recurrently found in cancer, with particular enrichment in chronic myelomonocytic leukaemia (CMML) where they are found in 60% of patients. TET2m has been implicated in immune cell dysregulation but its role in CMML pathogenesis remains unclear. Recent reports of hypermethylation of *cis*-regulatory elements for genes involved in innate immune cell activity in TET2m CHIP and MDS raises the possibility that TET2 mutation could epigenetically prime CMML haematopoietic stem/progenitor cells (HSPC) for myelomonocytic differentiation. I set out to characterise the genome regulatory landscape of TET2m CMML and in doing so uncovered a novel link between TET2m and emergency myelopoiesis (EM).

Methods: Gene regulatory changes in TET2m CMML were investigated using RNA and ATAC sequencing of bulk lineage-negative CD34+ HSPCs and paired single cell multiome analysis of bone marrow mononuclear cells. Imaging mass cytometry was used to validate cell-cell interactions in TET2m CMML. Response to M-CSF and CSF1R inhibition, using emactuzumab, was assessed using primary liquid culture of monocytes and HSPCs.

Results: Analysis of accessible chromatin in HSPCs subdivided TET2m CMML patients into three subgroups: erythroid-biased, myeloid-biased and balanced. Single nuclear multiomic analysis of patients within these subgroups found distinct changes to global haematopoietic gene regulatory networks. Erythroid-biased patients had normal activity of early erythroid TFs eg GATA1, but reduced activity in TFs controlling terminal erythroid differentiation eg NFIA, potentially explaining the increase in erythroid progenitors found in these patients. Myeloid-biased patients exhibited enhanced activity of CEBP family TFs, which promote myelomonocytic differentiation. Despite these global differences, HSCs from patients in both subgroups demonstrated increased CEBPB activity and reduced BACH2 activity. These are methylation sensitive TFs that act as a molecular switch controlling EM, where methylation promotes CEBPB binding at the expense of BACH2.

We reasoned that epigenetic priming might make HSPCs hypersensitive to cytokine signals that promote EM. Ligand-receptor interactions involving HSPCs were screened *in silico* and identified M-CSF as a potential trigger for EM in TET2m CMML. M-CSF and CSF1R expression on classical monocytes and HSPCs respectively were confirmed using imaging mass cytometry on bone marrow trephine samples. Furthermore, these populations were found to reside in closer proximity in TET2m CMML than in healthy donors. Primary culture of TET2m CMML HSPCs found 57% of CMML patients exhibited hypersensitivity to M-CSF, promoting HSPC proliferation and monocytic differentiation, which could be abrogated by emactuzumab.

Discussion: TET2 loss of function in CMML HSPCs creates a permissive epigenetic environment for myelomonocytic differentiation. This primes HSPCs for hyper-responsiveness to extrinsic cues from factors such as M-CSF that promote EM. Targeting of these pathways provides a novel therapeutic window for this orphan malignancy.

Contents

0.1 Lay abstract v
0.2 Scientific abstract vi

List of Figures **xv**

List of Tables **xix**

List of Abbreviations **xxi**

1 Introduction **1**
1.1 Chronic Myelomonocytic Leukaemia 1
1.1.1 Chronic myelomonocytic leukaemia (CMML) diagnosis and epidemiology 1
1.1.2 CMML Pathogenesis 3
1.1.3 Predicting prognosis in CMML 10
1.1.4 CMML treatment options and effectiveness 13
1.1.5 Ten-eleven translocation (TET) methylcytosine dioxygenase 2 19
1.1.6 TET2 mutation in cancer 28
1.1.7 Seed and Soil theory of cancer pathogenesis 31
1.1.8 Hypothesis 33
1.1.9 Specific aims 33

2 Materials and Methods **35**
2.1 Primary patient multiomic profiling 36
2.1.1 Human sample collection and biobanking 36
2.1.2 Targeted next generation sequencing 37
2.1.3 Thawing of cryopreserved MNCs 37
2.1.4 Fluorescent activated cell analysis and sorting (FACS) . . . 37
2.1.5 Bulk RNA sequencing 38
2.1.6 Bulk ATAC sequencing 40
2.1.7 Single nuclear RNA and single cell ATAC sequencing 43
2.1.8 Antibody conjugation and optimisation for imaging mass cytometry 44
2.1.9 Imaging mass cytometry 45

2.2 Primary patient cell culture assays 46
2.2.1 Primary monocyte culture assays 46
2.2.2 Primary HSPC proliferation assay 48
2.2.3 Primary HSPC survival assay 49
2.3 Bioinformatic analysis 52
2.3.1 Bulk ATAC-seq analysis 52
2.3.2 SmartSeq3 bulk RNA-seq analysis 53
2.3.3 Survival analysis 56
2.3.4 Single nucleotide polymorphism identification for demultiplexing individuals from 10X Multiome samples 56
2.3.5 Single nuclear RNA-seq and single cell ATAC-seq analysis . 58
2.3.6 Imaging mass cytometry analysis 69
2.3.7 Code availability 70

3 TET2 mutation primes CMML HSPCs for myelomonocytic differentiation by hijacking the emergency myelopoiesis pathway 71
3.1 Introduction 72
3.1.1 HSPC function in steady state haematopoiesis 72
3.1.2 Transcription factors coordinate HSPC differentiation 73
3.1.3 Emergency myelopoiesis and its triggers 75
3.1.4 HSPC dysfunction in CMML 76
3.1.5 Summary and plan for investigation 78
3.1.6 Aims 80
3.2 Results 80
3.2.1 HSPC populations in TET2m CMML 80
3.2.2 Patient and healthy donor characteristics of individuals included in multiomic bulk analysis 81
3.2.3 Spectrum of mutations in patient cohort 81
3.2.4 Chromatin accessibility of HSPCs separates TET2m CMML patients into three subgroups 83
3.2.5 Cluster 1 samples are enriched for SRSF2 mutations and cluster 3 samples for RAS pathway mutations 91
3.2.6 Multiomic single cell analysis reveals TET2m CMML HSCs are primed for myelomonocytic differentiation 91
3.2.7 Alterations to the gene regulatory landscape explain how different phenotypes develop within CMML 96
3.3 Discussion 99
3.3.1 Future work 103

4 Characterising changes to the extrinsic regulation of haematopoiesis in TET2m CMML 115
4.1 Introduction . . . 116
4.1.1 Extrinsic regulators of haematopoiesis . . . 116
4.1.2 Factors driving emergency myelopoiesis . . . 117
4.1.3 Structural changes to bone marrow architecture in CMML . 118
4.1.4 Non-haematopoietic stromal cells display altered abundance and function in CMML . . . 119
4.1.5 Altered cytokine signalling in CMML . . . 120
4.1.6 Mature myeloid cell populations undergo phenotypic changes in CMML . . . 122
4.1.7 Summary and plan for investigation . . . 124
4.1.8 Aims . . . 124
4.2 Results . . . 125
4.2.1 TET2 mutant CMML monocytes are transcriptionally and epigenetically distinct from healthy donor monocytes . . . 125
4.2.2 snRNA-seq facilitates identification of putative aberrant cell-cell interactions in TET2m CMML bone marrow . . . 130
4.2.3 Classical monocyte populations are expanded 10 fold in TET2m CMML bone marrow mononuclear cells and exhibit an aberrant immunophenotype . . . 132
4.2.4 TET2m CMML is associated with an increase in both classical monocyte and macrophage populations in the bone marrow . 135
4.2.5 HSPCs reside in close proximity to classical monocytes and macrophages in TET2m CMML bone marrow . . . 140
4.2.6 Multinichenet analysis identifies putative interactions that are predicted to promote HSPC dysfunction in TET2m CMML 140
4.3 Discussion . . . 150
4.3.1 Conclusion . . . 156

5 Ligand-receptor interactions can be targeted to inhibit cancer cell growth and function in TET2m CMML 159
5.1 Introduction . . . 160
5.1.1 Investigating the extrinsic regulators of haematopoiesis . . . 160
5.1.2 M-CSF and CSF1R in haematopoiesis . . . 161
5.1.3 Targeting M-CSF and CSF1R in cancer . . . 162
5.1.4 TGF-β signalling in haematopoiesis . . . 164
5.1.5 IL-1β signalling in haematopoiesis . . . 165
5.1.6 Summary and plan for investigation . . . 166

5.1.7 Aims . . . 166
5.2 Results . . . 167
5.2.1 TET2m CMML classical monocytes express CSF1R and NCAM1 . . . 167
5.2.2 CSF1R inhibition with emactuzumab leads to a dose dependent reduction in monocyte growth in the presence of M-CSF 168
5.2.3 Assays to test for dysregulation of HSPC response to extrinsic cytokine signalling in TET2m CMML . . . 170
5.2.4 IL-1β exerts a proliferative effect on HSPC culture and promotes myelomonocytic differentiation in TET2m CMML and healthy donors . . . 173
5.2.5 TGF-β causes dose dependent growth inhibition, suppression of myelomonocytic differentiation and enhanced erythropoiesis in TET2m CMML and healthy donor HSPCs . . . 173
5.2.6 Emactuzumab inhibits proliferation and monocyte differentiation of HSPCs in M-CSF responsive TET2m CMML patients 177
5.3 Discussion . . . 181
5.4 Conclusion . . . 186

6 Discussion 187
6.1 TET2 mutation creates a permissive epigenetic landscape for emergency myelopoiesis . . . 188
6.2 Expanded mature myeloid populations create a feed forward loop to enhance emergency myelopoiesis . . . 191
6.3 Extrinsic cytokine signalling can be targeted to correct HSPC dysfunction in TET2m CMML . . . 192
6.4 Implications and wider relevance . . . 194
6.5 Limitations . . . 195
6.5.1 Is deep molecular phenotyping of patient samples the correct approach? . . . 195
6.5.2 Limitations of chosen techniques . . . 197
6.6 Future Work . . . 198
6.6.1 How does TET2 dysfunction impact cell metabolism and is this a targetable vulnerability? . . . 198
6.6.2 What is the mechanism through which TET2 loss of function promotes C/EBP-β activity? . . . 199
6.6.3 Can the aberrant epigenetic priming be targeted to mitigate HSPC dysfunction in CMML? . . . 199
6.6.4 How does TET2 loss of function contribute to the pathogenicity of co-mutations? . . . 200
6.7 Concluding remarks . . . 201

Appendices

A Appendix **205**
A.1 Supplementary Tables 205
A.1.1 SmartSeq3 reagents and volumes 205
A.1.2 10X multiome reagents and volumes 205

References **211**

List of Figures

1.1 **TET enzyme expression throughout human tissues** 27
1.2 **Distribution of TET2 mutations in cancer** 30
1.3 **Summary of project aims** 32

2.1 **Bulk population sorting for RNA and ATAC sequencing** . . 40
2.2 **Sort purity assessment** 41
2.3 **Gating strategy to identify HSPC subpopulations and mature myeloid populations** 42
2.4 **Liquid culture output assessment for HSPC primary culture assays** 51
2.5 **ATAC-seq quality control assessment** 54
2.6 **ATAC-seq batch effect assessment** 55
2.7 **RNA-seq quality control and batch effect assessment** 57
2.8 **snRNA-seq quality control assessment** 59
2.9 **scATAC-seq quality control assessment** 61
2.10 **Comparison of marker gene expression using RNA or CellBender adjusted RNA counts** 62
2.11 **Discovery of ambiguous clusters containing cells with high levels of ambient RNA contamination** 64

3.1 **HSPC immunophenotype analysis** 82
3.2 **Somatic mutations of CMML patients used in bulk assays** . 84
3.3 **Multiomic analysis of HSPCs identifies three subgroups within TET2m CMML patients** 85
3.4 **Chromatin accessibility analysis identifies distinct phenotypic groups within TET2m CMML HSPCs** 86
3.5 **Motif enrichment analysis identifies phenotypic differences between TET2m CMML subgroups** 88
3.6 **RNA-seq analysis of HSPCs supports phenotypic division of TET2m CMML HSPCs discovered by ATAC-seq but variance between samples is more strongly driven by differences with healthy donors** 104

3.7 **Re-analysis of HSPC immunophenotype confirms phenotypic clusters identified by chromatin accessibility** 105
3.8 **Phenotypic clusters are associated with specific genotypes** . 106
3.9 **Overview of snRNA-seq and scATAC-seq multiomic profiling of TET2m CMML** 107
3.10 **snRNA-seq and scATAC-seq integration identifies multiomically defined cell states** 108
3.11 **HSPC subpopulations are well represented in all phenotypic groups** 109
3.12 **Gene expression analysis of HSC subcluster** 110
3.13 **Chromatin accessibility analysis of HSC subcluster** 111
3.14 **Exploring potential drivers of altered HSC function in CMML** 112
3.15 **Erythroid and myeloid-biased CMML are driven by distinct changes to the genomic regulation landscape** 113

4.1 **TET2m CMML monocytes have different chromatin accessibility to healthy donor monocytes but do not separate into HSPC defined phenotypic clusters** 127
4.2 **TET2m CMML monocytes have altered ligand expression to healthy donor monocytes** 128
4.3 **Subgroup analysis of mature myeloid cell populations improves accuracy of cell type annotation and identifies differences as well as similarities between CMML phenotypes** . . 133
4.4 **Changes in abundance and immunophenotype of mature myeloid populations in TET2m CMML** 134
4.5 **Overview of imaging mass cytometry experiment** 135
4.6 **Imaging mass cytometry confirms that TET2m CMML is associated with an expansion of classical monocyte and macrophage populations** 138
4.7 **Spatial analysis uncovers alterations to the cells within the HSPC niche in TET2m CMML** 139
4.8 **Multinichenet analysis identifies putative interactions predicted to support HSPC dysfunction in TET2m CMML** . . 144
4.9 **RNA expression of candidate ligands and receptors split by phenotypic group** 145
4.10 **HSPCs express multiple receptors of interest but there was no difference in expression between healthy donors and TET2m CMML** 147
4.11 **Ligand expression by different cell types within the bone marrow** 149

4.12 **Ligand expression in classical monocytes** 150

5.1 **NCAM1 and CSF1R are expressed by classical monocytes in TET2m CMML** 169
5.2 **Emactuzumab inhibits monocyte proliferation in a dose dependent manner in TET2m CMML and healthy donor monocytes** 171
5.3 **Primary HSPC culture to assess proliferation and survival in response to stimulation or inhibition of ligand-receptor interactions** 172
5.4 **IL-1β has a proliferative effect on HSPC culture and promotes myelomonocytic differentiation in TET2m CMML and healthy donors** 174
5.5 **TGF-β causes dose dependent growth inhibition, suppression of myelomonocytic differentiation and enhanced erythropoiesis in TET2m CMML and healthy donor HSPCs** 176
5.6 **The majority of TET2m CMML patients exhibit HSPC hypersensitivity to M-CSF, which is completely reversed by emactuzumab** 179
5.7 **M-CSF promotes survival of monocyte-primed HSPC at the expense of granulocyte-primed HSPC** 182

6.1 **Visual Summary - TET2 dysfunction epigenetically primes HSPCs for myelomonocytic differentiation using an emergency myelopoiesis-like response to extrinsic cytokine signalling** 194

List of Tables

1.1 Diagnostic criteria for chronic myelomonocytic leukaemia 4
1.2 Selected prognostic risk scores in chronic myelomonocytic leukaemia 12

2.1 Antibody panel for bulk RNA-seq and ATAC-seq sort 39
2.2 Antibody panel for single nuclear RNA-seq and single cell ATAC-seq sort 44
2.3 Imaging mass cytometry: Fluorophore conjugated antibody master mix 46
2.4 Imaging mass cytometry: Metal conjugated antibody master mix . 47
2.5 Antibody panel for monocyte immunophenotyping 48
2.6 Antibody panel for HSPC culture sort 49
2.7 Culture conditions for primary HSPC proliferation assays 50
2.8 Antibody panel for HSPC culture output analysis 50
2.9 Culture conditions for primary HSPC survival assays 52

3.1 TET2m CMML patient demographics 83

A.1 SmartSeq3 lysis buffer 206
A.2 SmartSeq3 reverse transcription mix 206
A.3 SmartSeq3 PCR mix 207
A.4 ATAC transposition buffer 207
A.5 ATAC index PCR mix 208
A.6 10X multiome lysis buffer 208
A.7 10X multiome wash buffer 209
A.8 10X multiome diluted nuclei buffer 209

List of Abbreviations

7-AAD . . . . 7-Aminoactinomycin D

AML . . . . . Acute myeloid leukaemia

AP-1 . . . . . Activating protein 1

ASXL1 . . . . Additional sex combs like (ASXL) transcriptional regulator 1

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

BM . . . . . Bone marrow

CCUS . . . . . Clonal cytopenia of uncertain significance

C/EBP . . . . CCAAT-enhancer binding proteins

CH . . . . . Clonal haematopoieisis

CHIP . . . . . Clonal haematopoieisis of indeterminate potential

CI . . . . . Confidence interval

CMML . . . . Chronic myelomonocytic leukaemia

CMP . . . . . Common myeloid progenitor

CPM . . . . . Counts per million

DNA . . . . . Deoxyribonucleic acid

DNMT . . . . DNA methyltransferase

EM . . . . . Emergency Myelopoiesis

FACS . . . . . Fluorescent-activated cell analysis and sorting

GSEA . . . . . Gene set enrichment analysis

GMP . . . . . Granulocyte-monocyte progenitor

HR . . . . . Hazard Ratio

HMA . . . . . Hypomethylating agent

HSC . . . . . Haematopoietic stem cell

HSPC . . . . . Haematopoietic stem and progenitor cell

ISO . . . . . International Organisation for Standardization

LMPP . . . . . Lymphoid-primed multipotent progenitor

M-CSF . . . . Macrophage-Colony Stimulating Factor

MD-CMML . Myelodysplastic CMML, white cell count <13x10^9/L

MDS . . . . . Myelodysplastic Syndrome, also known as Myelodysplastic Neoplasm (MDN)

MEP . . . . . Megakaryocyte-erythroid progenitor

MNCs . . . . . Mononuclear cells

MP-CMML . Myeloproliferative CMML, white cell count ≥13x10^9/L

MPN . . . . . Myeloproliferative neoplasms

MPP . . . . . Multiprotent progenitor

MSC . . . . . Mesenchymal stem cell

MSigDB . . . Molecular Signatures Database

PB . . . . . Peripheral blood

PBS . . . . . Phosphate buffered saline

PCA . . . . . Principal component analysis

PCR . . . . . Polymerase Chain Reaction

RNA-seq . . . Ribonucleic acid sequencing

SRSF2 . . . . . Serine and Arginine rich splicing factor 2

TET2 . . . . . Ten-eleven-translocation (TET) methylcytosine dioxygenase 2

TET2m . . . . TET2 mutant

TF . . . . . Transcription factor

TGF-β . . . . . Transforming growth factor-beta

TNF-α . . . . . Tumour necrosis factor-alpha

TP53 . . . . . Tumor protein 53

UMAP . . . . Uniform manifold approximation and projection

VAF . . . . . Variant allele frequency

VCF . . . . . Variant call format

WNN . . . . . Weighted nearest neighbour

1 Introduction

Contents

1.1 Chronic Myelomonocytic Leukaemia 1
1.1.1 Chronic myelomonocytic leukaemia (CMML) diagnosis and epidemiology 1
1.1.2 CMML Pathogenesis 3
1.1.3 Predicting prognosis in CMML 10
1.1.4 CMML treatment options and effectiveness 13
1.1.5 Ten-eleven translocation (TET) methylcytosine dioxygenase 2 19
1.1.6 TET2 mutation in cancer 28
1.1.7 Seed and Soil theory of cancer pathogenesis 31
1.1.8 Hypothesis 33
1.1.9 Specific aims 33

1.1 Chronic Myelomonocytic Leukaemia

1.1.1 Chronic myelomonocytic leukaemia (CMML) diagnosis and epidemiology

Chronic myelomonocytic leukaemia (CMML) is an orphan cancer with no disease specific treatments and is largely incurable. It is a myeloid malignancy that shares clinical features and pathobiology with both myelodysplastic syndrome (MDS) and myeloproliferative neoplasms (MPN). CMML originates from the altered function of

haematopoietic stem cells (HSCs), leading to excessive production of myelomonocytic cells, the most mature of which are neutrophils and monocytes. Patients often present indolently with non-specific symptoms such as tiredness or weight loss, which can be easy to attribute to symptoms of ageing or pre-morbid conditions. Routine peripheral blood analysis then finds an elevated monocyte count that persists for greater than three months, prompting further diagnostic workup. However, not all patients present in this way. CMML can present with dysplastic type symptoms, such as dyspnoea and haemorrhage secondary to cytopenia, or recurrent and atypical infections; or proliferative type symptoms, such as fever, night sweats, weight loss, cachexia, bone pain or splenomegaly[1]. At diagnosis, 25% of patients have anaemia requiring regular blood transfusions and 25% have an enlarged spleen[2], [3]. Some patients may present with symptoms of an associated disease, with 20-30% of CMML patients developing systemic inflammatory and autoimmune diseases (SIADs)[4], [5]. Rarely, CMML patients can present with symptoms from systemic mastocytosis[6], leukaemia cutis[7], or in transformation to blast phase CMML, which is classified as Acute Myeloid Leukaemia, myelodysplasia-related (AML-MR), because of shared clinical and pathological correlates with AML arising from MDS[8], [9].

CMML is the most common myelodysplastic/myeloproliferative neoplasm, a group of disparate myeloid malignancies that share certain clinical features with both MDS and MPN, that was created in the 2001 edition of the World Health Organisation (WHO) Classification of Tumours[10]. Prior to this CMML was a subtype of MDS. CMML is rare, with an age adjusted incidence of 0.4 in 100,000 people[11], [12]. Median age at diagnosis is 70-74 and the sex ratio is skewed, with the prevalence being double in men. However, women who are diagnosed with CMML have a worse prognosis[12], [13].

Diagnosis, according to the latest WHO and International Consensus Classification (ICC) 2022 criteria[14], is made through the analysis of peripheral blood and bone marrow samples. Peripheral blood for evidence of a persistent, absolute ($>0.5\text{x}10^9$/L) and relative ($\geq$10%) monocytosis. The WHO classification also recognises the presence of >94% classical monocytes on peripheral blood

flow cytometry, in the absence of active autoimmune disease and/or systemic inflammatory syndrome, as a novel supportive criteria. Neutrophil leukocytosis is very common, but less discriminating from alternative causes, so is not included in the diagnostic criteria. Bone marrow cytological and histological examination demonstrates dysplasia in ≥10% of cells of a myeloid lineage, and absence of disease defining molecular (eg. BCR-ABL) or pathological (>20% blasts) criteria for alternative haematological malignancies (Table 1.1)[9]. Both the WHO and ICC definitions recognise the importance of demonstrating clonality by targeted next generation sequencing (NGS) for mutations recurrently found in myeloid malignancies and for cytogenetic aberrations by karyotype analysis. The ICC criteria in particular stipulate that the variant allele frequency (VAF) of any mutation be ≥10% and that a clonal marker is required to make a diagnosis of CMML in a patient with a monocyte count of 0.5-1.0x10^9/L. Despite these precise diagnostic criteria, CMML, by its nature as an MDS/MPN crossover malignancy, can be difficult to distinguish from other MDS or MPN entities. Therefore, the careful consideration of all available pathology and molecular results by a multidisciplinary team is crucial for the accurate diagnosis of CMML. CMML can be sub-stratified by white cell count into myelodysplastic CMML (MD-CMML; WBC<13×10^9/L) and myeloproliferative CMML (MP-CMML; WBC≥13×10^9/L), which have molecular and prognostic significance discussed below. CMML is also staged by the proportion of blasts or promonocytes in the peripheral blood and bone marrow (see Table1.1).

1.1.2 CMML Pathogenesis

Cell of origin

The malignant CMML clone arises in multipotent haematopoietic progenitor cells, which sit at the apex of haematopoiesis, capable of self-renewal and multilineage differentiation. The most compelling evidence for this comes from xenograft transplantation models, where CMML patient cells are injected into severely immunodeficient mice and assessed for leukaemic engraftment. These experiments are technically challenging, and traditional murine models fail to sustain the CMML

Diagnostic criteria for Chronic Myelomonocytic Leukaemia
Prerequisite criteria:
1. Persistent absolute ($\geq 0.5\times10^9$/L) and relative ($\geq$10%) peripheral blood monocytosis 2. Blasts and promonocytes constitute <20% of the cells in the peripheral blood and bone marrow 3. Not meeting diagnostic criteria of chronic myeloid leukaemia or other myeloproliferative neoplasms 4. Not meeting diagnostic criteria of myeloid/lymphoid neoplasms with tyrosine kinase fusions
Supporting criteria:
1. Dysplasia involving $\geq$10% of cells from $\geq$1 myeloid lineage in bone marrow 2. Acquired clonal cytogenetic or molecular abnormality 3. >94% classical monocytes on peripheral blood flow cytometry in the absence of known active autoimmune diseases and/or systemic inflammatory syndromes
Requirements for diagnosis:
Pre-requisite criteria must be present in all cases - If monocytosis is $\geq 1\times10^9$/L: one or more supporting criteria must be met - If monocytosis is $\geq$0.5 and $<1\times10^9$/L: supporting criteria 1 and 2 must be met
Subtyping criteria:
- Myelodysplastic CMML (MD-CMML): WBC$<13\times10^9$/L - Myeloproliferative CMML (MP-CMML): WBC$\geq 13\times10^9$/L
Subgrouping criteria (based on percentage of blasts and promonocytes):
CMML-1: <5% in peripheral blood and <10% in bone marrow CMML-2: 5–19% in peripheral blood and 10-19% in bone marrow

Table 1.1: Diagnostic criteria for chronic myelomonocytic leukaemia - WHO 2022 classification, adapted from Khoury et al. Leukemia (2022)[9]

graft. However, engraftment is possible in NSG-SGM3 (NSGS) mice and the leukaemia initiating cells have been found to be within the $CD34^+$ $CD38^-$ multipotent haematopoietic stem and progenitor cells (HSPCs) and not within $CD34^+$ $CD38^+$ oligopotent HSPCs[15]. In a separate study of somatic mutations in single cell colonies grown from CMML HSPCs, 75% of CMML patients had evidence of mutations within $CD34^+$ $CD38^-$ multipotent HSPCs that were shared with $CD14^+$ monocytes. Furthermore, 44% also had mutations shared between $CD3^+$ T cells and $CD14^+$ monocytes, implying their origin within a multipotent HSPC[16]. Mutations

in TET2 and ASXL1 were frequently found to be the initiating mutation. Additional mutations were then serially acquired, leading to the establishment of progressive clonal dominance through the stages of myeloid differentiation. These findings have been confirmed using mitochondrial clone tracking in single cell RNA sequencing (scRNA-seq) and single cell DNA sequencing, demonstrating that clones originate within haematopoietic stem cells and rapidly achieve clonal dominance, resulting in oligoclonality even within multipotent HSPCs[17], [18]. This mirrors findings in MPN and MDS, where the leukaemia initiating cell has been identified as a haematopoietic stem cell (HSC)[19], [20].

Somatic mutations occur at a steady rate throughout life in all tissues of the body including blood[21]–[23]. These occur stochastically as a result of errors during DNA repair or replication. If they happen to occur within cancer driver genes, a gene where a mutation can provide a competitive advantage to the cell, then these cells will outcompete their wild type neighbours leading to clonal expansion of the original cell's progeny. When this occurs in an HSC of an otherwise healthy individual, it is called clonal haematopoiesis (CH). CH can be further subdivided into clonal cytopenia of undetermined significance (CCUS) where CH is associated with cytopenia or clonal haematopoiesis of indeterminate potential (CHIP) when blood counts are normal[24]–[27]. Driver genes discovered in CH are also recurrently found to be mutated in myeloid malignancy, with epigenetic regulators TET2, DNMT3A and ASXL1 being most frequently found. Mutations in DNA damage response proteins - TP53 and PPM1D; and spliceosome genes - SRSF2, U2AF1 and SF3B1; are also regularly found but at a lower prevalence. CH can develop at any age but incidence increases throughout life. Highly sensitive targeted error-corrected sequencing techniques have been developed, which are able to detect the presence of very small clones (VAF 0.0003%), and this has led to the discovery that CH is almost ubiquitous in individuals over 50 years old[28]. CH dramatically increases the relative risk of developing haematological malignancy, although the absolute risk remains small (0.5%-1% per year)[25]. CH clones increase in size over the life time of the individual and the dynamics of this clonal growth are dictated by

plethora factors including the specific mutation, the age at which the mutation was acquired and environmental factors such as the presence of chronic inflammation or smoking[22], [23], [29], [30]. As clones expand the risk of transformation to a myeloid neoplasm increases. The mechanisms and dynamics of this transformation are still being uncovered but several groups have used the UK Biobank resource to identify some specific risk factors for transformation[31], [32]. Clone size (VAF), specific genes/mutations, number of mutations, alterations to blood cell parameters such as red cell distribution width or cytopenia and age were all found to be predictive of transformation. Furthermore, certain combinations of mutations were enriched in those transforming to specific myeloid neoplasms, such as SF3B1 or TET2/SRSF2 comutation, which were associated with transformation to MDS, and JAK2 or CALR or MPL, which were associated with transformation to MPN[31].

Transformation of CH to CMML is thought to occur through clonal evolution, where CH clones acquire additional mutations resulting in functional alterations to clonal HSPCs leading to the development of CMML phenotype. Evidence for this is provided by deep mutational analysis using whole exome sequencing of patients with CMML, which demonstrates a high degree of overlap between mutations associated with CH and those found in CMML. ASXL1, EZH2 and SETBP1 mutations are associated with an increased number of additional mutations. However the great majority of CMML patients, who lack these specific alterations, continue to accumulate mutations at a similar rate to age matched controls even after being diagnosed with CMML[33]. The acquisition of mutations in CMML can therefore be modelled using a multi-hit stochastic model to predict the expected incidence of CMML given that on average 6-10 somatic exonic mutations are required for CMML to develop[33], [34]. This model accurately predicts the true age-adjusted incidence of CMML. This finding strongly suggests that CMML develops as a natural consequence of ageing through the same stochastic mechanism that causes CH. However, this deterministic view of CMML developing in elderly people through nothing more than bad luck does not mean the development of CMML, or any other haematological malignancy, is inevitable. In fact, the vast majority of patients with

CH never progress to haematological malignancy. Therefore, if it were possible to adjust known risk factors for CH progression then many cases of haematological malignancy could be prevented.

Despite the excitement surrounding the potential to instigate a screening programme for CH, questions remain about what can be done to mitigate an individual's risk of progression to haematological malignancy. The key requirement to unlocking this opportunity is to understand the drivers of CH clonal expansion, and identify therapies that can prevent it. This is a rapidly evolving field, and some mutation specific drivers have been identified, such as in JAK2 mutations and TP53 mutations. JAK2 V617F mutation leads to constitutive signalling from several haematopoietic growth factor receptors leading to proliferation and clonal expansion, just as seen in MPN, where it is found in over half the patients[35]. TP53 facilitates sensing of DNA damage and can instigate cell cycle arrest and apoptosis when it occurs. Loss of function mutations therefore provide a survival advantage if severe DNA damage is caused by irradiation or cytotoxic chemotherapy[36]. To explore the drivers of clonal expansion more broadly, *Fabre et al* conducted a longitudinal study of CH and found that most mutations convey a cell-intrinsic, mutation specific fitness advantage that is steady over time[22]. Some expand at a slower rate, such as DNMT3A and TP53, of 5% per year, whereas others such as SRSF2, PTPN11 and U2AF1 expand at an increased rate of 15-20% per year. This suggests inherent differences in the strength of the fitness advantage conveyed by mutations in different genes. However, it was also possible to identify a small number of clones that did not grow at a steady rate. Only 58% of JAK2 V617F clones expanded uniformly over time and TET2 clones grew more rapidly in elderly individuals, indicating that context is important, and cell-extrinsic factors can play a role in speeding up or slowing down clone growth. Further evidence for the role of extrinsic regulation of clonal expansion has come from work prompted by the unexpected observation that patients with CH have an increased risk of atherosclerotic cardiovascular disease[25], [37]. Murine models of CH have implicated pro-inflammatory changes to clonal mature myeloid cells that promote immune cell

infiltration, and atherosclerotic plaque size expansion, via increased IL-6 and IL-1β secretion[37]–[39]. This same mechanism has also been implicated in promoting CH clonal expansion, where cell-intrinsic effects of the clonal mutation protect clonal HSPCs from the cell-extrinsic inflammatory stress created by clonal mature myeloid cells. Multiple *in vivo* animal disease models involving TET2, DNMT3A and C/EBPα knockout (KO) [40]–[46] have demonstrated the importance of this interaction to promoting clonal HSPC expansion, as well as several studies using mutation aware single cell RNA sequencing in humans with mutation in DNMT3A, TET2 and TP53[47], [48]. Given that CMML has been demonstrated to develop by the accumulation of stochastically acquired somatic mutations within a single clone[33], [34], the same mechanism as CH, it is highly likely that the interactions between cell-intrinsic fitness advantage and cell-extrinsic inflammatory pressure, which determine clonal expansion in CH also contribute to CMML pathogenesis.

Genetic determinants of CMML

CMML is caused by the progression of CH through the accumulation of multiple additional somatic mutations within clonal haematopoietic stem cells[16]. The original CH clone will acquire on average 2 additional driver mutations prior to diagnosis[34]. These mutations most frequently occur within epigenetic regulators (TET2 (56%), ASXL1 (41%)), spliceosome apparatus (SRSF2 (42%), SF3B1(7%), U2AF1 (7%), ZRSR2(6%)), RAS pathway (NRAS (15%), KRAS (9%), CBL(14%)), or oncogenic transcription factors (RUNX1 (15%), SETBP1 (9%))[49]. Mutations in these genes however are not specific to CMML, in fact mutations in all of these genes are also recurrently found in MDS and AML[50]. Despite significant efforts it remains unknown why certain patients develop a particular myeloid malignancy. The sequence in which different mutations are acquired is undoubtedly important, with the order of acquisition of TET2 and JAK2 mutations having a profound effect on myeloproliferative disease biology[51]. Recently, the epigenetic state of HSPCs at the time the initial driver mutation is acquired, in this case Tet2 KO, was found to correlate with the function of the clone[52]. Therefore, the ultimate disease

phenotype emerging from CH is determined by multiple factors, including the order in which they were acquired, the interaction between the effects of each mutation and the intrinsic properties of the cell of origin. In CMML, clone tracing via single cell culture or clonal reconstruction using variant allele frequencies, have identified TET2, SRSF2 and ASXL1 as the most frequent founder mutations, and RAS pathway mutations being secondary events[16], [34]. These findings suggest that epigenetic dysfunction is required to prime the clone prior to acquiring the proliferative effect of secondary mutations and the subsequent transformation into CMML.

Certain genetic aberrations can be pathognomonic for a particular disease phenotype. In MDS for instance, SF3B1 mutation causing MDS with ring sideroblasts (MDS-SF3B1), deletion of the long arm of chromosome 5 (MDS-5q), and biallelic TP53 inactivation (MDS-biTP53) have all been recognised as distinct entities within the latest WHO classification of myeloid malignancies[9]. However, although CMML phenotypes are highly heterogeneous between patients, very few molecular correlates have been identified. Several studies have found RAS pathway mutations to be enriched within proliferative CMML and a causal link via the novel RAS-KMT2A-PLK1 axis has been established[34], [53], [54]. JAK2 V617F mutations are also associated with a myeloproliferative CMML phenotype, with elevated haematocrit and platelet count also being seen[55]. In the absence of ASXL1 mutation, TET2 mutations confer a favourable prognosis and are associated with a non-proliferative, dysplastic phenotype[34], [49]. Frameshift and nonsense mutations in ASXL1 have also been found to be associated with inferior prognosis[1], [56]. These data strongly suggest that genotype can dictate CMML biology, but the complex interactions between multiple mutations can make it difficult to accurately identify clinical or biological correlates. Further evidence for interactions between mutations comes from data on their co-occurrence. EZH2 loss of function mutations, which are found in 5-7% of CMML patients, almost exclusively co-occur with ASXL1[57]. Mutations in SRSF2 however, are mutually exclusive with EZH2, but positively associated with TET2 co-occurrence[58]. Understanding the interplay between these different genetic factors will help us divide CMML further into

subcategories with shared biology, which will ultimately aid the identification of targeted therapies and improve patient care.

Structural chromosomal abnormalities are highly prevalent within cancer, including haematological malignancies[59]. However cytogenetic abnormalities are only found in 20-30% of patients with CMML[13], [60], [61]. In a study of 403 CMML patients, *Wassie et al* found the most common abnormalities are +8 (23%), -Y (20%), -7/7q$^-$ (14%), 20q$^-$ (8%), +21 (8%) and der(3q) (8%), with complex karyotype ($\geq$ 3 cytogenetic abnormalities) found in 11% [61]. Although cytogenetic abnormalities are infrequent, they can be highly prognostic, and cytogenetic risk has been incorporated into all CMML specific prognostic models[1], [13], [56], [62]–[65]. Isolated -Y and der(3q) are low risk, monosomal and complex karyotypes are high risk and all other abnormalities are intermediate risk[61].

1.1.3 Predicting prognosis in CMML

Newly diagnosed patients with CMML have a median survival of 3 years but this is highly variable between individual patients[13]. Building accurate predictive models of prognosis in CMML is important for informing treatment decisions and patient choice. They may also be used to identify inherent differences in disease biology between patient subgroups. Multiple prognostic models have been created, with the majority aiming to predict prognosis at the time of diagnosis[1], [13], [56], [62]–[64], although some have tried to identify patients that would most benefit from particular treatments, such as allogeneic stem cell transplantation[65]. These models have become increasingly sophisticated over time, with the latest incorporating cytogenetic and molecular criteria to more accurately separate patients into different prognostic groups. Selected studies are summarised in Table 1.2. Although scoring criteria are different for each of these models, there are a few recurring themes:

Increased age at diagnosis is associated with worse prognosis in CMML. Increased age is associated with increased co-morbidity, frailty and side effects from treatment[66]. However, ageing is also linked to increased inflammation (termed

inflammaging)[67] and altered HSC function, therefore physiological changes associated with age may be biologically important for worsening CMML prognosis[68]. However, from a molecular perspective, increased age is positively associated with the incidence of TET2 mutation, which confers a favourable prognosis in some models[49]. Age is also a significant determinant of whether a patient is eligible for an allogeneic haematopoietic stem cell transplant, the only curative treatment of CMML[65]. These data therefore add confounding to the interpretation of age in CMML prognostic models.

Cytopenia, or evidence of bone marrow failure, is also frequently seen in prognostic scoring criteria. Like age, cytopenia will increase the symptomatic burden of CMML, compounding co-morbidity or frailty of the individual. However, it also indicates the degree to which the CMML clone is perturbing normal haematopoiesis, which could be used as a surrogate of disease severity. The mechanism by which the CMML clone impairs haematopoiesis is not currently well understood. In MDS multiple mechanisms have been uncovered, such as excessive cleavage of GATA-1, a transcription factor (TF) essential for erthropoiesis, by caspase 3 due to defective protection of GATA-1 by its chaperone, heat shock protein 70 (HSP70)[69]. Haploinsufficiency of ribosomal protein RPS14 has been implicated as the cause of dyserythropoiesis in 5q- syndrome[70] and alternative 3' splicing of ABCB7 causes iron accumulation in ring sideroblasts in MDS-SF3B1[71]. These disparate mechanisms tend to be genotype specific and similar mechanistic links in CMML have not yet been discovered.

Cytogenetic abnormalities, other than -Y and der(3q), are associated with inferior prognosis[61]. For cytogenetic abnormalities to be detected, the clone must tolerate a structural aberration to chromatin whilst dividing and thereby have dysfunctional DNA damage response mechanisms. This will facilitate the acquisition of additional cytogenetic aberrations and somatic mutations, thereby increasing the risk of transformation to AML.

Myeloproliferative phenotype, can be associated with greater constitutional symptom burden due to excessive inflammatory cytokine release and metabolic

Model	Size (n)	Significant variables in model	Median survival by category (months)	Transformation to AML
MDAPS[64]	213	Hb<12g/L; Immature myeloid cells in PB; Lymphocyte count >2.5x10^9/L; BM blasts >10%	Low - 24; Int-1 - 15; Int-2 - 8; High - 5;	19% transformed to AML. Median time 7 months
CPSS[13]	578	FAB phenotype; WHO CMML stage; Cytogenetics; RBC transfusion dependence;	Low - 72; Int-1 - 31; Int-2; 13; High - 5;	13%, 29%, 60% and 73% at 5 years respectively
GFM[56]	312	Age > 65yrs; WBC >15x10^9/L; Hb <10g/dL; Platelets ≤100x10^9/L; ASXL1 mutation;	Low - Not reached; Int - 38.5; High - 14.4;	AML-free survival 56, 27.4 and 9.2 months respectively
Mayo Molecular Model[1]	466	Monocyte count >10x10^9/L; Blasts in PB; Hb <10g/dL; Platelets ≤100x10^9/L; Frameshift or nonsense ASXL1 mutations;	Low - 97; Int-1 - 59; Int-2 - 31; High - 16;	16% transformed to AML by median follow up of 23 months
CPSS-Mol[63]	214	Genetic risk - cytogenetics + ASXL1, NRAS, SETBP1 and RUNX1 mutations; BM blasts ≥5%; WBC ≥15x10^9/L; RBC transfusion dependence;	Low - Not reached; Int-1 - 64; Int-2 - 37; High - 18;	48 months cumulative incidence of AML - 0%, 3%, 21% and 48% respectively

Table 1.2: Selected prognostic risk scores in chronic myelomonocytic leukaemia - Adapted from *Patnaik and Tefferi, AJH(2024)*[73]. Abbreviations: PB, peripheral blood; BM, bone marrow;MDAPS, MD Anderson Prognostic Score; CPSS, CMML-specific prognostic scoring system; FAB, French Americaln British; WHO, World Health Organisation; GFM, Groupe Francophone des Myelodysplasies;

imbalance leading to cachexia. However there are likely to be other important biological implications as to why an elevated leukocyte count conveys a poor prognosis, such as the association with RAS pathway mutations and ASXL1 mutations in some studies[72].

Transformation to Acute Myeloid Leukaemia

20-30% of patients will transform from CMML to AML, which is usually refractory to chemotherapy, and has a prognosis of just 6 months[57]. No current treatments, other than allogeneic haematopoietic stem cell transplantation (HSCT) can alleviate

this risk, although getting patients to transplant is challenging due to the poor response to chemotherapy and hypomethylating agents seen in this cohort. The mechanisms by which patients transform to AML are not fully understood. Adverse cytogenetics[13], [61], peripheral blood blasts[1], ASXL1, RUNX1, NRAS, SETBP1, DNMT3A ad NPM1 mutations[1], [72], [74] are all predictive of transformation. These risk factors provide some insight into the mechanism of transformation. Presence of an increased number of blasts in the bone marrow or blasts entering the peripheral circulation are likely to indicate that differentiation block is emerging and the blast forming clone will inevitably go on to expand into classically defined AML once the blast percentage is above 20%. Similar reasoning has led to discussions about the most appropriate cut off for blast percentage in *de novo* AML, where the latest WHO classification has dispensed with a fixed blast percentage threshold in the presence of certain pathognomonic genomic alterations[9]. As discussed above, the presence of cytogenetic abnormalities is an indicator of dysfunction in DNA damage response pathways and therefore this environment will also facilitate clonal evolution leading to transformation to AML. This does not preclude that the altered balance of gene expression within chromosomes, which have greater or fewer copies of genes as a result of the structural alteration, may also have a profound impact on transformation. In the same way, mutations in specific genes, such as NPM1, an AML defining mutation, or RUNX1 will increase the risk of transformation. Whole exon sequencing of CMML patients have on average 19 somatic mutations, compared to 29 mutations in patients after transformation from CMML to AML[34]. In a longitudinal study of CMML, serial samples were able to identify the presence of clonal evolution secondary to acquiring new cytogenetic abnormalities in 24% and novel somatic mutations in 50%[73].

1.1.4 CMML treatment options and effectiveness

Despite high symptomatic burden, 'real world' observational studies have found that only half of the patients with CMML receive any disease targeting treatment, although practice is heterogeneous between countries and institutions[3], [75]–

[78]. Patients not receiving disease modifying therapy receive 'best supportive care'. Cytopenia is alleviated by blood product transfusions when required and infection is rapidly treated with appropriate antibiotics. Multiple therapies have been investigated to improve patient quality of life through combating cytopenias. Anaemia can be targeted with erythropoiesis stimulating agents (ESAs) to mitigate the need for blood transfusion. In the largest study of the effectiveness of ESAs in CMML, response was seen in 64%, but this is often short lived, with a median duration of response being 8 months[79]. Serum erythropoietin levels were able to predict the likelihood of response. Transforming growth factor-β (TGF-β) superfamily ligand traps, sotatercept and luspatercept, have shown great promise in low risk MDS for reducing transfusion dependency, in particular MDS-SF3B1, and multiple phase II and III studies have been reported[80]–[82]. Although patients with CMML were permitted in these studies, the overall number of CMML patients included is not reported and presumed to be very small. It therefore remains to be seen whether this is an effective treatment strategy in this patient group. Thrombocytopaenia can be severe in CMML, resulting in some patients requiring regular prophylactic platelet transfusions to prevent bleeding. Thrombopoietin receptor agonist eltromobopag has been assessed in a recent multicenter phase II trial in CMML[83], which found overall response rate (ORR) was 72%. Responses were rapid, within a median of 14 days, but short-lived (median 2.6 months), making it a potential option to treat a patient who is undergoing a planned procedure, but not for chronic prevention of bleeding in a thrombocytopenic patient.

Hydroxycarbamide is the most well-established disease targeting treatment, which can reduce constitutional symptoms, leukocytosis and splenomegaly. Median duration of response is 24 months. It is not thought to prolong survival, however there are no prospective data comparing with best supportive care, so its overall effect on survival is unknown[84]. Hydroxycarbamide also suppresses normal haematopoiesis and can therefore worsen anaemia and thrombocytopenia, precluding its use in many patients.

DNA methyltransferase inhibitors (DNMTi), also known as hypomethylating agents (HMA) are licensed for treatment of advanced stage CMML with myelodysplastic phenotype. Azacitidine (AZA) gained a license following a subgroup analysis of the CALGB[85] and AZA-001[86] registry studies that compared AZA with best supportive care in patients with MDS or AML. These studies contained a combined total of just 25 patients with CMML. In patients that respond, it can reduce their requirement for blood transfusion and there was initially hope that it improved patient survival. However, the DACOTA trial, a more recent phase III study (n=170) comparing decitabine (DAC) to hydroxycarbamide in high risk proliferative CMML, found no overall survival advantage. Response rate was 63% for DAC vs 35% for hydroxycarbamide, AML transformation was reduced in DAC treated patients (HR 0.62, 95% CI 0.41-0.94; p=0.005), but mortality rate was increased(HR 1.55, 95% CI 0.82–2.9; p=0.04), with no difference in median overall survival (18.4 vs 21.9 months, p=0.67)[87]. No head to head trials have been performed, but retrospective data have found that AZA and DAC perform similarly; ORR 50% and CR 10-20% [75], [88]. Oral decitabine also performs similarly to the intravenous preparation [89]. Detailed analysis of effects of HMAs in clinically defined responders and non-responders found that responders have a profound reduction in genome-wide methylation and corresponding altered gene expression, whereas non-responders had no change to either assay when reviewing serial samples before and after HMA exposure. Furthermore, responders had no reduction in the CMML clone size, indicating that the clinical response to HMAs is an epigenetic phenomenon[90]. Mutant HSCs transiently regain haematopoietic function to normalise peripheral blood counts, but the absence of a clonal response helps explain why disease responses tend to be short lived[18]. Therefore their use remains controversial in the field, but in the absence of good alternatives they remain first line treatment for high risk myelodysplastic CMML[91].

The only curative option for patients with CMML is allogeneic haematopoietic stem cell transplantation, but this is highly toxic and only a minority (<10%) of patients are young enough and fit enough to tolerate the procedure. Response

rates in CMML patients following HSCT are 17-50% and therapy related mortality 12-52% [88]. A recent large retrospective analysis of international registry data found that lower risk CMML patients had worse outcome following HSCT (20% (with HSCT) vs 42% (without HSCT) 5yr survival) but higher risk CMML patients had an improved outcome (27% vs 15% 5yr survival)[92]. CMML is therefore a cancer in great need of new treatment options.

Novel therapies being investigated in CMML

Venetoclax, an inhibitor of the anti-apoptotic protein BCL-2, in combination with azacitidine, has emerged as a transformational treatment for elderly patients with AML in recent years but has not been prospectively assessed in CMML[93]. However, responses in retrospective case series and preclinical data have been modest. In a case series of 27 patients, of which 74% had failed HMA therapy, ORR to venetoclax in combination with HMA or cytotoxic chemotherapy was 67% with complete response in 4%[94]. Preclinical data in CMML has mixed results to support the use of venetoclax in CMML. Monocytes from patients with CMML are resistant to apoptosis due to dependence on myeloid cell leukaemia-1 (MCL-1), rather than BCL-2, potentially explaining why outcomes in CMML appear inferior to those in AML[95]. Conversely, Bcl-2 was increased in haematopoietic progenitors when Tet2 was knocked out in a murine model of CMML[41]. This may indicate that venetoclax will be effective in the TET2 mutant subset of CMML patients, but prospective trials on CMML patients without previous exposure to HMA are required to make these assessments.

CMML HSPCs are hypersensitive to granulocyte-macrophage colony stimulating factor (GM-CSF), particularly those with RAS pathway gain of function mutations[96], [97]. JAK/STAT signal transduction is integral to the activation of this pathway downstream of multiple important haematopoietic growth factor receptors and therefore JAK/STAT inhibition is an attractive target in CMML. Ruxolitinib, a JAK1/2 inhibitor, has been studied in a phase I/II clinical trial in CMML. ORR was 38%, with 43% having a reduction in splenomegaly. Encouragingly, grade

3/4 anaemia and thrombocytopenia developed in 10% and 6% respectively and response was seen in patients without mutation in RAS pathway genes as well as in MD-CMML patients[98]. Pacritinib, a JAK1 inhibitor that also has activity against macrophage colony stimulating factor (M-CSF) and IRAK1, has shown encouraging pre-clinical results in CMML patient xenografts[99] and synergy when used in combination with an HMA *in vitro*[100]. GM-CSF has also been targeted directly using a neutralising antibody lenzilumab. Following on from a positive phase I study, which found good tolerability in 15 patients with CMML[101], a phase II study (PREACH-M) using lenzilumab in combination with HMA in RAS-pathway mutant CMML has achieved 86% complete response rate with many patients also demonstrating a clonal response, which is usually not seen with HMA therapy alone[102]. These data are hugely encouraging and support the use of HMA based combination therapy in CMML. MEK is a downstream target kinase of the RAS/MAPK signalling pathway and inhibition of MEK has been effective at perturbing oncogenic RAS/MAPK signalling in multiple cancers[103]. MEK inhibitor cobimetinib is currently in phase II trial for RAS pathway mutant CMML, with an interim report demonstrating an ORR of 50% and satisfactory safety profile[104]. Tipifarnib, a farnesyl transferase inhibitor, is another novel therapy targeting the RAS/MAPK pathway in CMML. All isoforms of RAS are farnesyl transferase substrates, with HRAS being particularly dependent on farnesylation for its activity[105]. However, tipifarnib may also have activity targeting CXCL12/CXCR4 interactions, which are important in bone marrow homing of myeloid cells and retention of HSPCs. A phase II single arm open label study of tipifarnib in 37 CMML patients found it was well tolerated, with an ORR was 22%, and patients with KRAS and/or NRAS mutations had a superior response rate to wild type patients[106].

Multiple other therapies are currently being trialled in CMML targeting disparate mechanisms. High dose vitamin C is being tested in TET2 mutant CMML (NCT03418038). TET2 loss of function mutations are the most frequent somatic mutation found in CMML, and TET function can be partially restored by vitamin

C, which enhances residual TET2 and TET3 activity. This has been shown to reverse aberrant self-renewal of Tet2 deficient HSPCs in pre-clinical models[107]. Tagraxofusp, a novel diphtheria toxin conjugated fusion protein targeting CD123 (IL-3 receptor) designed to target leukaemic stem cells, is in a phase I trial in CMML. An interim report demonstrated significant toxicity but impressive response, with 11% achieving bone marrow morphologic complete response and all 11 patients having a splenic response in a heavily pre-treated cohort of CMML patients[106].

Identifying novel therapeutic targets by studying inflammatory signalling in CMML

As outlined above, the treatment landscape for CMML is rapidly evolving, with some promising early stage trials underway. However many of these novel therapies are targeting the RAS/MAPK pathway. This approach in MPN, where aberrant JAK/STAT signalling has been identified as being fundamental to pathogenesis, has thus far failed to produce significant reductions in clone size let alone a cure[108]. It is therefore likely that effects of single agent therapy are going to be modest and combining therapies in a targeted manner will be required to improve patient outcomes. To do this we need more treatments targeting distinct pathways and the best way to discover these are to improve our understanding of the fundamental biology of CMML. CMML is in essence a malignancy caused by HSPC dysfunction. CMML HSPCs preferentially differentiate to myelomonocytic lineages and exhibit enhanced stem cell properties through serial replating analysis[16]. Part of these aberrant functions will be attributable to the cell-intrinsic effects of the mutations, epigenetic and cytogenetic changes. However, as seen in CH, the way in which these cell-intrinsic effects interact with environmental cell-extrinsic factors ultimately dictates changes to function or phenotype. In CMML the bone marrow becomes packed with dysplastic, dysfunctional myeloid cells, leading to immunodeficiency and cytopenia. The large numbers of pro-inflammatory myelomonocytic cells release cytokines such as IL-8, IL-6, tumour necrosis factor alpha (TNF-α) and M-CSF creating a chronic inflammatory environment[109]–[111]. There is emerging evidence that clonal monocyte derived factors can influence the behaviour of CMML cells, with

cytokine-like protein 1 (CYTL1) aiding CMML monocytes to escape apoptosis, and macrophage migration inhibitory factor (MIF) promoting monocytic differentiation of myeloid progenitors[95], [112]. Clonal myeloid derived suppressor cells have also been implicated in suppressing monocyte differentiation by releasing alpha-defensins and suppressing wild type, but not clonal HSPC growth, by secreting IL-8[111], [113]. Myeloid progenitor cells in CMML exhibit hypersensitivity to GM-CSF, enhancing HSPC clonogenicity and activating JAK-STAT signalling[97]. Furthermore, unsorted CMML mononuclear cells can spontaneously form colonies in methylcellulose in the absence of exogenous growth factor, which can be inhibited by the addition of IL-10[114]. These data highlight cytokine signalling as a rich source of putative therapeutic targets in CMML. In this thesis I will outline how I have explored this further in TET2 mutant CMML using a multi-omic approach to discover a number of putative targets and validate M-CSF/CSF1R axis as a novel therapeutic target in CMML.

1.1.5 Ten-eleven translocation (TET) methylcytosine dioxygenase 2

TET2 as an epigenetic regulator

Ten-eleven translocation (TET) dioxygenases are a family of three, Fe^{2+} and α-ketoglutarate dependent enzymes that oxidise 5-methylcytosine (5-mC) leading to its active and passive demethylation. The catalytic activity of TET enzymes convert 5-mC to 5-hydroxymethylcytosine (5-hmC), which can then be removed by two active demethylation pathways[115]. Either through iterative conversion by TET to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), both of which are recognised by thymine DNA glycosylase (TDG), and removed by the TDG-mediated base excision repair pathway[116], [117]. Alternatively 5-hmC can be actively removed by the activation-induced cytidine deaminase(AID)-APOBEC DNA repair pathway, via conversion to 5-hydroxymethyluracil (5-hmU) followed by TDG or single-stranded selective monofunctional uracil DNA glycosylase (SMUG1)-mediated base excision repair[118], [119]. 5-hmC is not recognised by DNA methyltransferase 1 (DNMT1),

which is responsible for copying methylation patterns on to daughter strands during DNA replication, and therefore 5-hmC is passively removed during cell division[120].

DNA methylation is essential for mammalian embryonic development, with early lethality observed in mice that lack DNMT enzymes[121]. DNMTs are a family of three enzymes made up of DNMT1, responsible for maintaining methylation on replicating DNA, DNMT3a and DNMT3b, which perform *de novo* methylation of DNA[120], [122]. DNA methylation, with the exception of embryonic stem cells, occurs almost exclusively at CG dinucleotides (CpG). CpG are distributed sparsely throughout the genome and represent around 1% of total DNA bases. Although there is variation between different cell types, around 70-80% of CpG are methylated, particularly at gene bodies and intergenic regions[123], [124]. Approximately 2% of CpG are found clustered together in short (approximately 1kb) regions of increased CpG density, termed CpG islands (CGI). CGI are hypomethylated compared to average genomic methylation and roughly half are found in close proximity to transcription start sites (TSS) within the annotated promoter regions of 60-70% of genes (pCGI)[125], [126]. The other half of CGI, called orphan CGI (oCGI), are found distal to promoter regions and co-localise at primed enhancer elements[127], [128]. CGI are therefore highly enriched at *cis*-regulatory elements (CRE) and when hypomethylated are permissive of gene expression. Evidence for this comes from observations that identified hypermethylated CGI as facilitating gene silencing in X chromosome inactivation[129], gene imprinting[130] and suppression of transposable elements[131]. The mechanism whereby CGI avoid *de novo* methylation is not fully understood, with some speculating that CRE of genes that are highly expressed in early embryogenesis are shielded from contact with DNMT enzymes by proteins within the transcriptional apparatus when a wave of *de novo* methylation occurs as cells differentiate from embryonic stem cells after implantation[132], [133]. More recently, G4-quadruplexes, a rare secondary DNA structure enriched in telomeres and promoters, have also been implicated in helping to shield pCGI from methylation[134], [135]. Further work is required in this area to help understand how CGI are maintained in a hypomethylated state on a genome wide scale.

Although methylation of pCGI results in gene repression, the converse is not always true. pCGI are located in the promoter region of all house keeping genes, but 30-40% of pCGI are found at cell type specific genes. These pCGI remain hypomethylated throughout different cell types where these genes have very low or no detectable expression[136]. *In vivo* footprinting analysis of three cell type specific pCGI associated genes, c-JUN, PCNA and HBA1, found promoter regions to be constitutively bound by multiple ubiquitous TFs, with little variation between cell types. This suggests that these genes are actively repressed by TF occupancy and that additional cell type specific regulatory signals are required to initiate gene transcription[137]. A genome wide meta-analysis of CGI uncovered an additional mechanism of regulation of these regions, which relies on differences in density of CpG methylation in the CpG shores immediately adjacent to pCGI. Greater difference in methylation density between CGI shores and pCGI correlates with gene expression, with steep gradients, called CpG ravines, being found at promoter sites in highly active housekeeper genes[138].

As described above, CGI methylation is critical for regulation of gene expression, however recent whole genome methylation sequencing experiments have found that the greatest differences in methylation between cell types are found in CpG outside of CGI[139]. Using 205 samples of 39 different cell types *Loyfer et al.* performed an organism wide review of genomic methylation differences between human cell types. They found that 97% of differentially methylated CpG were unmethylated in a single cell type and methylated in all other cell types. Only 1.7-2.7% of these unmethylated CpG were found within CGI. Cell type specific unmethylated CpG were found in close proximity to cell type specific genes and to co-localise with H3K27ac, H3K4me1, DNaseI hypersensitive sites and chromHMM enhancer annotations[140]. These findings indicate that cell type specific unmethylated CpG are located at sites of active enhancers in open chromatin, facilitating the expression of cell type specific genes. The small fraction of differentially methylated CpG that were hypermethylated (3%) in one cell type compared to the others, were enriched for CpG residing in CGI (38%), H3K27me3 and polycomb. TF motif enrichment of

these loci found enrichment for CTCF binding sites, indicating that methylation can be deployed to alter CTCF binding and therein the 3D structure of DNA loops formed at these loci altering gene regulation within topologically associated domains[139]. Novel tools making use of catalytically inactive Cas9 (dCas9) fused to TET or DNMT3 enzymes have been used to confirm these findings, as well as the importance of methylation at promoter and enhancer CRE in controlling gene expression[141]–[143].

Whole genome bisulfite sequencing has facilitated the discovery of differences between cell types in health and disease at unprecedented resolution. However, it is important to also consider the mechanisms by which 5-mC can alter gene expression. Methylation of DNA alters its biomechanical properties and alters the ability of proteins, such as TFs and other transcriptional apparatus, to bind to DNA[144], [145]. TFs possess motif-binding domains that recognise specific DNA sequences (TF binding sites, TFBS). High throughput screening of the effect of methylation within TFBS has been performed on a library of 542 TFs[146]. 34% of TFs had enhanced binding to methylated TFBS, 23% had reduced binding to methylated TFBS, 11% had mixed or minimal difference to binding methylated TFBS and 33% of TFs did not recognise a motif containing CpG[146]. This surprising finding, that the majority of TFs are sensitive to methylation when binding DNA, and that more TFs had enhanced binding to methylated TFBS than unmethylated TFBS, has disrupted the previously held consensus that methylation always inhibited TF binding. The other best described mechanism through which methylation alters gene expression is through the function of methyl-CpG-binding domain proteins (MBD). Methylated CpG are specifically recognised and targeted by MBD[147], which are able to recruit histone modifiers, transcriptional co-repressors and promote local chromatin condensation, each of which can influence local gene expression[148].

The interplay between DNMT and TET enzyme activity dictates the position of 5-mC throughout the genome. This in turn influences chromatin organisation and gene expression. But what role does TET2 specifically have in regulating gene expression? Unlike TET1 and TET3, TET2 lacks a zinc finger CXXC DNA binding

domain and therefore relies on binding to other proteins in order to be brought into close proximity with CpG[149]. TET2 can interact with CXXC4 (also known as IDAX, the ancestral TET2 CXXC domain that split from TET2 during vertebrate evolution) and CXXC5, which accommodate binding to unmethylated CpG[150], [151]. However, CXXC4 may act more as a negative regulator of TET2 activity rather than a facilitator, as CXXC4 binding leads to caspase-mediated degradation of TET2[150]. TET1 and TET3 bind preferentially to TSS, and are therefore critical for maintaining low 5mC levels within promoter regions[152], [153]. Conversely, TET2 binds to enhancer regions and rarely to CGI or promoters[154]–[156]. In order for TET2 to be recruited to enhancer regions it must associate with a vast array of proteins, which possess their own specific DNA binding domains, such as TFs. TET2 has been found within the oestrogen receptor (ER) complex in ER+ breast cancer[154], [157], to be physically interacting with Wilms tumour protein (WT1) in AML[158] and to bind to SMAD nuclear interacting protein 1 (SNIP1) as a bridge to c-MYC target genes[159]. The inherent DNA binding properties of the TF to which TET2 associates with dictate its role. TET2 can perform a negative regulatory role when co-localising with TF that preferentially bind methylated-CpG such as WT1[158]. In the case of pioneer TFs, like C/EBP-β and OCT4, TET2 can facilitate opening up of chromatin to allow additional TFs to be recruited to the region[146], [160]. Chromatin can also be altered via a non-enzymatic function of TET2, its ability to bind to and therefore recruits other epigenetic modifiers, such as O-linked β-N-acetylglucosamine transferase (OGT)[161] and histone deacetylase 2 (HDAC2)[162].

TET2 function in haematopoiesis

TET enzymes are not uniformly expressed throughout human tissues (see Figure 1.1). TET2 and TET3 are both highly expressed within the bone marrow, implying that they play a key role in haematopoiesis[163]. Mulitple *Tet2* KO models have been developed in transgenic mice using haematopoietic specific Cre recombinase promoters[164]–[167]. *Tet2* KO in HSCs leads to the expansion of the HSPC

pool, with enhanced repopulation capacity in competitive transplantation models, myelomonocytic proliferation and extramedullary haematopoiesis, phenocopying many of the features of CMML[164]–[168]. Aged Tet2 KO mice also transformed to acute leukaemia, with a longer latency to transformation in Tet2 haploinsufficiency compared to homozygous KO[167]. Similarly in human *in vitro* assays, knockdown of *TET2* in multipotent cord blood HSPCs led to skewing of differentiation towards granulomonocytic cells opposed to erythroid cells and impaired teminal erythroid differentiation[16], [169]. Interestingly impairing the AID/APOBEC base excision repair pathway by *Aid* KO has a similar effect on HSC differentiation to *Tet2* KO in mice. *Aid* KO leads to myelomonocytic differentiation bias and impaired erythropoiesis, with enhanced *Cebpa* and reduced *Gata1* expression in granulomonocytic progenitors (GMP) and megakaryocyte-erythroid progenitors (MEP) respectively[170]. However, HSCs did not develop a competitive repopulation advantage, nor did *Aid/Tet2* or *Aid/Flt3-ITD* double KO result in myeloid malignancy. Therefore, impairing active AID/APOBEC associated demethylation perturbs normal erythropoiesis and facilitates myelomonocytic differentiation but the enhanced stem cell fitness and leukaemogenesis observed in *TET2* KO are acquired through either a different mechanism, such as altered chromatin, or methylation at different loci to those altered by *Aid* KO. Similar findings were reported when comparing a catalytically inactive Tet2 with Tet2 KO, where catalytical dysfunction led to dysplastic morphology and proliferative myelomonocytic differentiation bias, but did not display enhanced HSC fitness through serial replating assays[171]. Different to Aid KO however, 79% of mice with Tet2 catalytic dysfunction developed transplantable myeloid malignancies, suggesting both catalytic and non-catalytic functions are important for TET2's role as a tumour suppressor in haematological malignancy.

CpG methylation, similar to chromatin accessibility and gene expression, undergo stereotyped alterations through haematopoietic stem cell differentiation and differences between cell populations can be used to accurately identify cell types[172]–[176]. Haematopoietic cell type specific areas of open chromatin are hypomethylated in

HSCs[172]. As HSCs differentiate into more mature progeny, cell type specific open chromatin areas become methylated in regions that are not related to the cell's differentiation trajectory. *De novo* methylation, with associated closing of chromatin and repression of inappropriate gene expression, appears fundamental to HSC differentiation[172]. However, traditional analysis of bulk sorted populations of immunophenotypically defined cells lacks sensitivity to uncover heterogeneity within the HSC population. Single cell analysis using a random fluorescent reporter to track single HSC clones *in vivo* found that HSCs are epigenetically primed for particular functions, such as lineage-specific differentiation or proliferation, rather than quiescence[177]. This creates an environment where a relatively small proportion of primed HSC clones produce the majority of blood cells whilst other HSCs are held in reserve[178]. Therefore, observed changes identified in bulk level methylation analysis between HSC and MPP populations may not be solely due to *de novo* methylation, but rather a lack of sensitivity to detect the primed HSC clones that become enriched as HSCs differentiate. Changes in methylation and chromatin accessibility are highly correlated with observed function, far more so than transcriptional differences[177]. Therefore, epigenetic priming of HSCs is deterministic of function, and it is of no surprise that alterations to epigenetic machinery, such as loss of TET2 function, can have profound effects on HSC function.

The molecular consequences of TET2 loss have been studied by measuring 5-hmC on purified bulk haematopoietic populations from healthy donors and *TET2* KO human HSPCs[179]. In healthy donors 5-hmC levels positively correlated with transcript expression and chromatin accessibility and are typically found distributed throughout gene bodies, extending up to 10kb away from TSS and transcription termination site. This suggests that TET enzymes may interact directly with transcriptional machinery during active transcription, as well as at enhancer sites. 5-hmC levels were depleted by *TET2* KO, particularly at erythroid cell specific genes. Motif enrichment analysis of annotated enhancers found reduced 5-hmC at erythroid TF motifs such as GATA1 and GATA2 but increased 5-hmC at myeloid TF motifs such as C/EBP family members, perhaps indicating that TET2 is required to

maintain hypomethylation at erythroid TF target sites, but TET1 or TET3 provide this function for myeloid TF target sites[179]. An alternative explanation is that 5-hmC is a highly dynamic epigenetic mark, which is enriched at genes that are actively being transcribed rather than it being permissive of transcription in the same way 5-mC functions. This is perhaps a more likely explanation because previous work on TF preference for methylated or unmethylated TFBS, found that GATA1 and GATA2 did not possess CpG in their TFBS and C/EBP family preferentially bound methylated TFBS[146]. Furthermore, individuals with TET2 CHIP and clonal cytopenia of uncertain significance (CCUS) have increased methylation at C/EBP family and ETS family TFBS, TFs responsible for myeloid differentiation, in keeping with predicted preferences for methylated TFBS[180].

In summary, TET2 is particularly important for maintaining hypomethylated enhancers in HSPC, and thereby supports balanced haematopoietic differentiation. Disruption of TET2 catalytic function or TET2 KO leads to myelomonocytic bias in HSC differentiation by inhibiting the expression of erythroid TFs and enhancing the function of myeloid TFs. The mechanism by which it exerts this function needs to be fully elucidated, but *in vitro* treatment of TET2 KO cells with either ascorbate or 5-azacitidine, reverses these effects, suggesting that TET2 is specifically required to remove methylation from certain genomic loci, which otherwise are responsible for this phenotypic change[179]. Greater understanding of these targets would inform our understanding of the epigenetic regulation of haematopoiesis and potentially identify treatment targets for TET2 mutant haematological malignancies.

TET2 as a regulator of inflammation

TET2 exerts disparate effects on different haematopoietic cells during inflammation. TET2 acts as a negative regulator of inflammation in mature myeloid cells. *Tet2* expression is increased in macrophages exposed to lipopolysaccharide (LPS), a glycolipid expressed on the surface of gram negative bacteria[181]. Following inflammatory stimuli, TET2 plays a critical role in recruiting HDAC2 to switch off IL-6 expression during inflammation resolution in dendritic cells and macrophages[162].

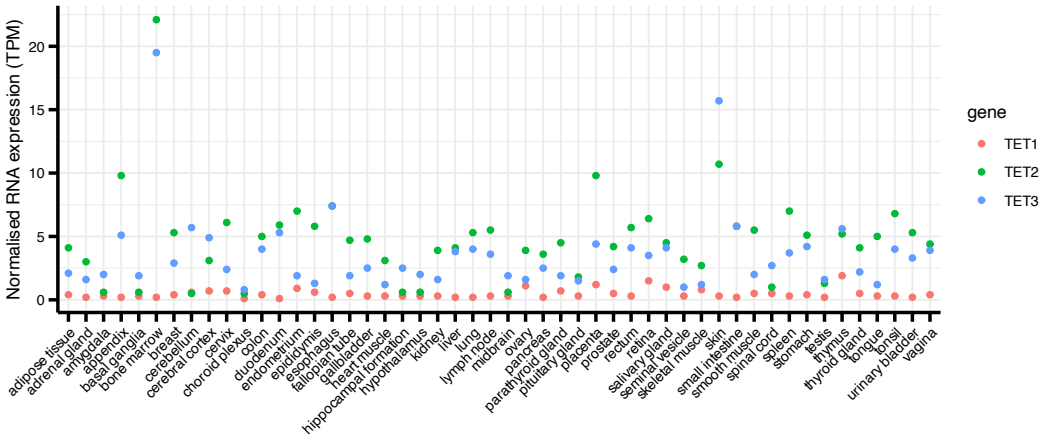


Figure 1.1: TET enzyme expression throughout human tissues - Normalised consensus RNA expression of TET1-3 displayed as transcripts per million (TPM) from bulk analysis of human tissue samples. Data downloaded from the Human Protein Atlas and includes data from HPA and GTEx projects[163].

Therefore, TET2 loss of function is associated with increased inflammation, which can contribute to the pathogenesis of multiple diseases. *Tet2* haploinsufficiency has been identified as a cause of excess IL-1β secretion by macrophages, which is a key driver of atherosclerosis and heart failure in TET2m CHIP[38], [182]. Conversely, the immunosuppressive activity of TET2 can also contribute to disease. Tumour associated macrophages require TET2 to maintain an immunosuppressive phenotype, and *Tet2* deletion results in a pro-inflammatory phenotypical switch, interferon secretion, adaptive immune cell activation and reduced tumour burden in a murine model of melanoma[183].

Deregulation of mature myeloid cell inflammatory signalling can also have direct consequences on HSPC function. For instance, increased IL-6 expression, driven by aberrant macrophage activity secondary to bacterial gut translocation in *Tet2* KO mice, has been implicated in driving pre-leukaemic myeloproliferation[42]. Furthermore, TET2 loss of function can protect HSPCs from apoptosis induced by exposure to the pro-inflammatory cytokines IL-6 and TNFα in the context of acute inflammation[40], [41]. TET2 loss of function also provides a proliferative advantage during chronic inflammation. Murine and *in vitro* models of chronic inflammation

created using repeated IL-1 exposure or toll like receptor (TLR) activation causes impaired self-renewal in TET2 wild type HSPCs, but not in *TET2* KO HSPCs[184]–[186]. This phenomenon has also been demonstrated using single cell molecular analysis of individuals with TET2 mutant CHIP where it was found that mutant HSCs down-regulate interferon and canonical NFκB immune response pathways to gain a competitive advantage over wild type HSCs within an inflammatory niche[48]. Therefore, TET2 loss of function can alter both the cell-intrinsic and cell-extrinsic factors supporting clonal expansion. This has direct relevance to TET2 mutant cancer, where inflammation is a well recognised hallmark[187], and in particular CMML, which is characterised by the dramatic expansion of mature myeloid cell populations, and pro-inflammatory cytokine release[109]. This raises the possibility that inflammatory signals from mature clonal myeloid cells are fundamental to the pathogenesis of TET2m CMML by directly promoting HSPC proliferation and myelomonocytic differentiation.

1.1.6 TET2 mutation in cancer

TET2 loss of function mutations are frequently found in many different types of cancer, with particular enrichment in haematological malignancies (Figure 1.2A). cBioPortal is an online repository of genomic data from a large number of cancer studies[188]. 75% of TET2 mutations on the cBioPortal database are truncating mutations, which will impair mRNA transcription and result in nonsense mediated decay of mRNA, reducing protein translation[189]. 18% are missense mutations, concentrated around either the N- or C-terminal end of the catalytic domain. Characterisation of multiple missense point mutations within the catalytic domain have been performed confirming that a single amino acid substitution is sufficient to disrupt the catalytic function of TET2[171], [190], which is most likely the consequence of these missense mutations given their distribution. Likewise splice site mutations found concentrated in Exons 4-10 are likely to affect catalytic function or translation. TET2 amplification is very infrequently found in cancer registries, representing 0.04% of cases within the GENIE consortium database[191]. Therefore

TET2 functions as a tumour suppressor, likely via multiple mechanisms as discussed above. Some reports have also implicated TET2 as having a role in DNA damage response, and consequently have an increased sensitivity to PARP inhibition[192], [193]. TET2's function in active demethylation helps mitigate the inherent risk of mutagenesis posed by 5-mC, but it remains controversial as to whether TET2m enhances mutagenesis. Some reports have found a small increase in con-current mutations in TET2m patients, but there are confounding explanations that could account for this, such as increased patient age[49], [194].

TET2 mutation in CMML

TET2 loss of function mutations are the most frequently found genetic abnormality in CMML (60%) and bi-allelic loss of TET2 function, defined as compound heterozygous mutations with VAF $>$ 55% or a homozygous mutation with VAF$>$55% and either loss of 4q22 or copy neutral loss of heterozygosity by uniparental disomy, is enriched in CMML compared to other myeloid malignancies[13], [34], [49], [56], [195]–[199]. This has led to the suggestion that possessing multiple TET2 mutations could be incorporated into the diagnostic classification of CMML[200]. In the largest study of the clinical correlates of TET2m in myeloid malignancy, *Awawda et al* found that biallelic loss of TET2 function was associated with truncating mutations, more elderly individuals, monocytosis (including in non-CMML myeloid neoplasms), low blast count and normal karyotype[195]. TET2m in CMML has been associated with myelodysplastic phenotype (although this has not been consistently found), increased age, co-mutation with SRSF2 and ZRSR2, and a favourable prognosis in the absence of ASXL1 mutation[34], [49]. In the largest study of TET2m in CMML, 75% of patients had truncating mutations, which were uniformly distributed throughout the coding sequence, similar to mutations reported in other malignancies[49]. TET2 mutations were associated with high clonal burden, with a median variant allele frequency of 44%.

The molecular consequences of TET2 loss of function have been the focus of multiple studies. As expected, TET2m CMML is associated with loss of 5-hmC

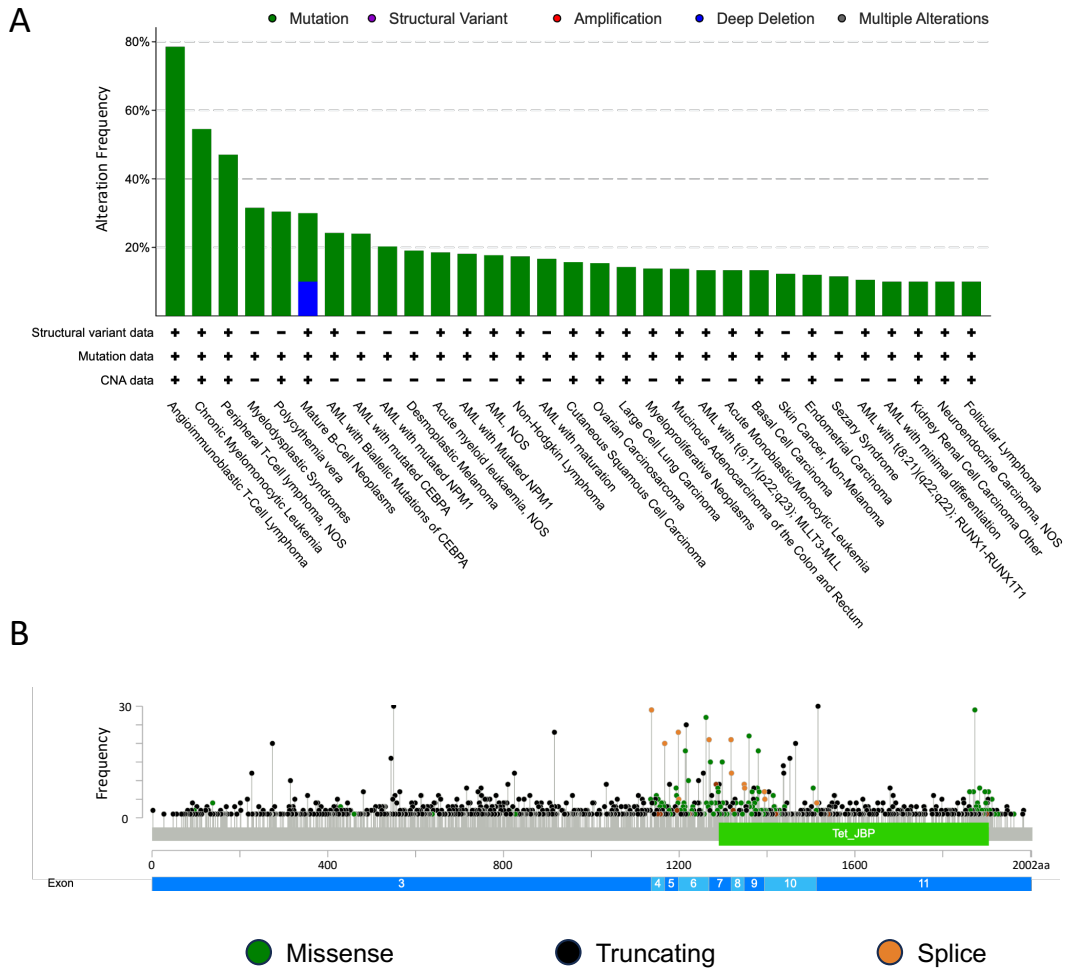


Figure 1.2: Distribution of TET2 mutations in cancer - (A) Frequency of TET2 mutation by cancer type. Bars are coloured by the type of genetic alteration in TET2 detected within each sample. All cancer types with >10 samples in cBioPortal and >10% TET2 alteration frequency are displayed. Data type availability for samples included in the analysis are displayed in rows. CNA, copy number alteration. (B) Lollipop plot displaying frequency of 2453 mutations from 1726 samples from all cancer types in cBioPortal database. Tet_JBP is the catalytic domain of TET2. Graphs were created using the cBioPortal online platform[188].

at CpG and increased 5-mC[190], [201], [202]. Early studies were hampered by the use of low resolution microarray (22k probes) or pyrosequencing technologies. Although resolution was sufficient to find that TET2m CMML patients clustered separately to TET2 wild type CMML patients, it was not sufficient to confidently identify pathways that were differentially targeted[201]. A more recent study using a microarray with 450k probes found hypermethylation at probes in proximity to genes enriched for gene ontology terms such as immune response, leukocyte activation, proliferation or migration, and monocyte or granulocyte chemotaxis; similar to terms enriched within hypermethylated regions in TET2m CHIP[180], [203]. Interestingly, TET2 wild type CMML patients with high risk features, such as abnormal karyotype or advanced stage, also have an altered methylation profile distinct from TET2m patients. These differences are potentially driven by increased metabolic or replicative stress associated with these adverse features[201], [203]. Specific methylation changes at individual loci have been implicated in predicting patient response to HMA therapy, suggesting that methylation in these regions is essential to CMML pathogenesis[204], [205]. Similarly, methylation change at specific promoter regions has also been identified as a regulator of individual genes that could influence CMML phenotype, such as *TIF1G* and *miR150*[206], [207].

1.1.7 Seed and Soil theory of cancer pathogenesis

The Seed and Soil theorem of cancer was first proposed by Stephen Paget in 1889 to explain the non-random distribution of distant metastatic sites in breast cancer[208]. He hypothesised that the environment had to be correct for cancer to successfully seed to particular tissues. Following the advent of genomics, our understanding of the molecular basis to this phenomenon has dramatically improved and the various factors involved with shaping tumour growth within the tumour micro-environment have been elucidated for many different cancers. Recurrent themes are apparent, known as the hallmarks of cancer, the malignant clone must possess a cell-intrinsic competitive advantage over it's wild type counterparts, through increased proliferation, resistance to cell death, resistance to growth

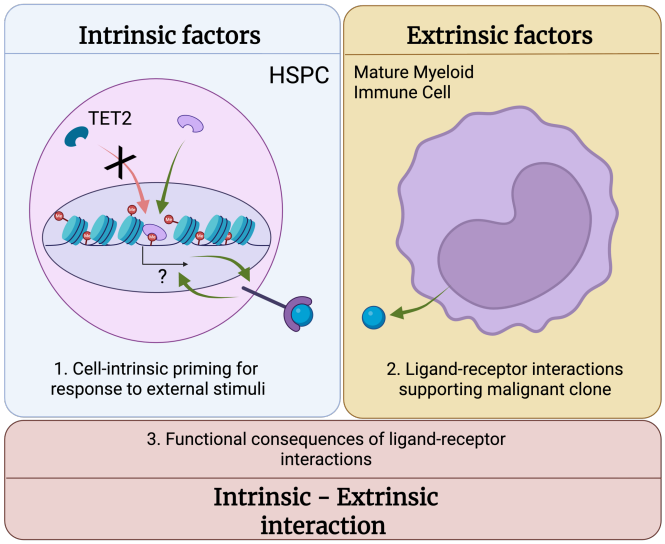


Figure 1.3: Summary of project aims - During this thesis I will explore the cell-intrinsic and cell-extrinsic drivers of TET2 mutant CMML. Each specific aim will be covered by a separate results chapter, culminating in the identification of a novel therapeutic target for CMML.

suppressors, evading immune surveillance and enabling replicative immortality. The tumour micro-environment provides cell-extrinsic factors to support tumour growth by the recruitment of immune cells to create a supportive inflammatory milieu, release growth factors that encourage tumour cell invasion and angiogenesis[187]. Although these hallmarks were devised largely through the study of solid organ malignancies, they remain relevant to haematopoietic malignancies, which are uniquely positioned to be able to directly shape the immune micro-environment to enhance tumour growth.

As outlined above, TET2m results in cell-intrinsic changes to HSPCs, providing protection from the repressive effects of inflammatory signalling, as well as promoting pro-inflammatory function of mature myeloid cells. Cell-extrinsic signalling between these two populations therefore has the potential to be fundamental to CMML pathogenesis. I chose to investigate this axis further in the work presented in this thesis, with a view of identifying a potentially targetable interaction that may provide a novel treatment option for this orphan malignancy.

1.1.8 Hypothesis

TET2 loss of function mutations in HSCs confer a competitive clonal advantage, particularly in the context of localised inflammatory signalling. During clonal expansion TET2 mutant HSPCs become epigenetically primed for myelomonocytic differentiation. Acquisition of additional driver mutations leads to clonal haematopoiesis progressing to CMML, which in turn leads to the aberrant proliferation of mature myelomonocytic cells within the bone marrow, fuelling further inflammatory drive. My hypothesis is that this creates a feed-forward loop, where clonal HSPCs require continuous interaction with clonal mature myelomonocytic cells to sustain their growth, and that by carefully characterising these interactions we may find novel therapeutic targets to help combat TET2 mutant CMML (summarised in Figure 1.3).

1.1.9 Specific aims

Aims:

1. Characterise cell-intrinsic epigenetic priming of TET2 mutant CMML HSPCs
2. Identify cell-extrinsic ligand-receptor interactions capable of supporting aberrant HSPC function in TET2 mutant CMML
3. Assess the functional importance of these interactions and screen them for potential as a therapeutic target in TET2 mutant CMML

2

Materials and Methods

Contents

2.1 Primary patient multiomic profiling 36
2.1.1 Human sample collection and biobanking 36
2.1.2 Targeted next generation sequencing 37
2.1.3 Thawing of cryopreserved MNCs 37
2.1.4 Fluorescent activated cell analysis and sorting (FACS) . 37
2.1.5 Bulk RNA sequencing 38
2.1.6 Bulk ATAC sequencing 40
2.1.7 Single nuclear RNA and single cell ATAC sequencing . . 43
2.1.8 Antibody conjugation and optimisation for imaging mass cytometry 44
2.1.9 Imaging mass cytometry 45
2.2 Primary patient cell culture assays 46
2.2.1 Primary monocyte culture assays 46
2.2.2 Primary HSPC proliferation assay 48
2.2.3 Primary HSPC survival assay 49
2.3 Bioinformatic analysis 52
2.3.1 Bulk ATAC-seq analysis 52
2.3.2 SmartSeq3 bulk RNA-seq analysis 53
2.3.3 Survival analysis 56
2.3.4 Single nucleotide polymorphism identification for demultiplexing individuals from 10X Multiome samples 56
2.3.5 Single nuclear RNA-seq and single cell ATAC-seq analysis 58
2.3.6 Imaging mass cytometry analysis 69
2.3.7 Code availability 70

2.1 Primary patient multiomic profiling

2.1.1 Human sample collection and biobanking

Samples were collected and processed for cryopreservation with the help of Nawshad Hayder, Cristina Tormo Garcia and other members of the Haematopoietic Stem Cell Biology Group

Bone marrow aspirate and peripheral blood samples from CMML patients and healthy donors were collected into or supplied by HaemBio biobank (UK REC: 17/SC/0572) and the INForMeD study (UK REC: 16/LO/1376). Healthy donor samples came from three potential sources: Bone marrow samples were either taken from individuals with suspected haematological malignancy, which was subsequently excluded and samples confirmed as normal by cytology/histopathology, or from femoral heads removed for elective total hip replacement. Peripheral blood samples were G-CSF mobilised samples from NHSBT where either excess cells were collected or the sample was no longer required for allogeneic stem cell transplantation. Written informed consent was obtained prior to sample collection and use in research in accordance with the Declaration of Helsinki. Total mononuclear cells (MNC) were separated from peripheral blood and bone marrow using Ficoll-Paque Premium (Merck, GE17-5442-03) gradient centrifugation according to manufacturer instructions and cryopreserved in FBS supplemented with 10% DMSO (D2650-100ml).

Formalin fixed paraffin embedded bone marrow trephine sections were obtained from excess clinical samples stored within the Oxford Radcliffe Biobank (UK REC: 19/SC/0173). Consent for this use was obtained within the INForMeD study (UK REC: 16/LO/1376) for which this has been explicitly approved. Sections were prepared by the Oxford Centre for Histopathology Research with paired haematoxylin and eosin stained sections and images.

2.1.2 Targeted next generation sequencing

Variant calling and classification was performed by Dr Angela Hamblin

Samples with unknown genotype had DNA extracted using DNeasy Blood and Tissue Kit (Qiagen, 69506) from total MNCs. Libraries were prepared using a custom designed, ISO accredited, Roche SeqCap EZ HyperCap platform with baits designed to capture 67 genes recurrently mutated in myeloid disease. Samples were sequenced on an Illumina MiSeq instrument and analysed using an in-house pipeline by Dr Angela Hamblin[209]. Clinical significance of detected variants were graded by Dr Angela Hamblin and visually assessed on Integrated Genomics Viewer. Variants are only reported when likely to be of clinical significance. Historical samples were either assayed as above or sequenced using a TruSeq custom amplicon panel (Illumina) for 32 recurrently mutated genes in myeloid neoplasms as previously published[209]. These samples had variant calling performed either by GATK[210] or somatic variant caller (Illumina).

2.1.3 Thawing of cryopreserved MNCs

Cryopreserved MNCs were thawed rapidly at 37°C in a water bath. 1ml thaw media (Iscove's Modified Dulbecco's Medium without phenol red (Gibco, 11500556) supplemented with 20% FBS and 1mg/ml DNase I (Merck, 260913)) was added dropwise to the cryovial and transferred to a 50ml tube where additional thaw media was added slowly to make 30ml. The cell suspension was centrifuged at 300G for 10 minutes at 4°C and supernatant removed. Cells were resuspending in the appropriate buffer or media for the downstream assay.

2.1.4 Fluorescent activated cell analysis and sorting (FACS)

Cells were washed in FACS buffer (PBS supplemented with 5% FBS ± 2mM EDTA) and counted using trypan blue either by haemocytometer or LUNA II (Logos Biosystems) automated cell counter. Prior to staining, cells were incubated for 10 minutes on ice with Fc Block buffer (PBS supplemented with 20% TruStain FcX (BioLegend, 422302)). Fc blocked cells were stained on ice for 20-30 minutes in the

dark using antibodies outlined in the relevant Table of this methods chapter. Cells were washed in FACS buffer and centrifuged at 400G for 5 minutes. Supernatant was removed and cells were resuspended in ice cold FACS buffer before being passed through a 35µm cell strainer (Falcon, 352235). 7-amino-actinomycin (7AAD, eBioscience, 00-6993-50) was added at 1:50-1:200 to assess cell viability. Cell sorting was performed using a Becton Dickinson (BD Biosciences) FACSAria Fusion instrument and a 100nm nozzle on high purity mode. Sort strategy for bulk RNA and ATAC sequencing is outlined in Figure 2.1. Sorted cell purity was confirmed by re-running a test population of sorted cells to confirm marker expression was in keeping with gating strategy used for the initial sort (see Figure 2.2). Data were collected during these sorts for additional markers of mature myeloid cell population and for subtyping HSPC populations (Table 2.1 and Figure 2.3). The HSPC marker panel and gating strategy is well established and has been validated in our lab and others [176], [211], [212]. The mature myeloid panel is less rigorous due to constraints of the instrument used. Therefore pDC and cDC populations are likely to be less pure, and it is not possible to distinguish cDC1 from cDC2.

2.1.5 Bulk RNA sequencing

Performed with assistance from Dr Christina Simoglou Karali

2000 monocytes per sample were sorted directly into RLT buffer and RNA was extracted using the RNeasy micro kit (Qiagen, 74004) in accordance with manufacturer's instructions. Initial attempts to prepare libraries from 200 lineage negative $CD34^+$ HSPCs sorted directly into SmartSeq3 lysis buffer failed due to RNA degradation and therefore RNA and gDNA were extracted from HSPC cell pellets using the AllPrep micro kit (Qiagen, 80284). Pellets were lysed as they thawed in RLT plus buffer. RNA and gDNA extraction was performed according to manufacturer's instructions and eluted in 10µl of RNase-free water. RNA quality was assessed using RNA 6000 pico kit on the bioanalyzer (Agilent, 5067-1513). RNA integrity number (RIN) was consistently good, although the RIN was reduced when RNA concentration was below the sensitivity range recommended by the

Table 2.1: Antibody panel for bulk RNA-seq and ATAC-seq sort

Fluorophore	Target	Clone	Supplier	Titration (μl per 100μl)
BV421	CD90	5E10	Biolegend (328122)	3.33
BV570	HLA-DR	L243	Biolegend (307638)	2
BV650	CD45RA	HI100	Biolegend (304136)	2
BV711	CD115	9-4D2-1E4	Biolegend (347318)	5
BV785	CD16	3G8	Biolegend (302046)	0.8
BB700	CD11c	3.9	BD Biosciences (748270)	0.83
PE	CD14	M5E2	Biolegend (982508)	3.33
PE-Texas Red	CD38	HIT2	Invitrogen (2208339)	5
PE/Cy7	CD123	6H6	eBioscience (25-1239-42)	1.67
APC	CD64	_10.1	Biolegend (305014)	3.33
APC-eFluor 780	CD34	4H11	eBioscience (47-0349-42)	0.666666667
Lineage cocktail				
FITC	CD3	SK7 (Leu-4)	Biolegend (345763)	2
FITC	CD8a	RPA-T8	Biolegend (301006)	1
FITC	CD10	HI10a	Biolegend (312208)	3.33
FITC	CD19	HIB19	Invitrogen (11-0199-42)	2
FITC	CD20	2H7	Biolegend (302304)	0.67
FITC	CD235a	10F7MN	eBioscience (11-9886-42)	2

manufacturer. Full length mRNA libraries were created using SmartSeq3 chemistry as previously published[213]. Due to increased starting RNA quantity, all reactions were performed in twice the volume of the published protocol. 4µl of RNA was used as input for each reverse transcription reaction replacing RT-PCR grade water in the lysis buffer. Samples received 12 to 18 cycles of preamplification PCR and cDNA libraries were cleaned up using AMPure XP beads (0.6:1 ratio, Beckman Coulter, A63881). 200pg of cDNA was tagmented using Nextera XT library preparation kit (illumina, FC-131-1096) at reduced volumes (2μl cDNA, 1.8μl tagmentation buffer and 0.2μl amplicon tagmentation mix). Libraries were indexed using Nextera XT Index Kit v2 (illumina, FC-131-2004) primers and 12 or 14 cycles of indexing PCR for HSPC and monocyte libraries respectively. Indexed sequencing libraries were cleaned up using AMPure XP beads (0.6:1 ratio). A complete list of reagents and volumes are provided in the Supplementary Tables

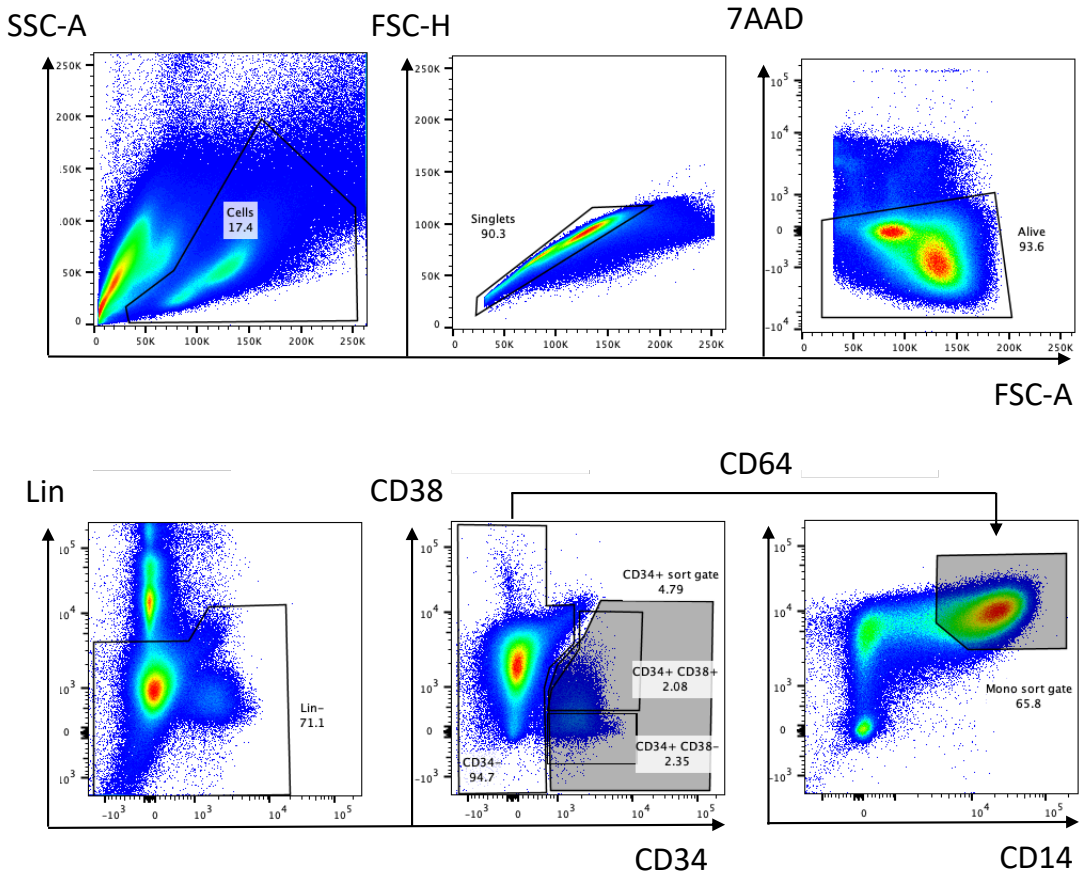


Figure 2.1: Sorting strategy for bulk RNA and ATAC sequencing - Lin^- $CD34^+$ HSPCs and $CD14^+$ $CD64^+$ monocytes were sorted either directly into RLT buffer or PBS

A.1, A.2 and A.3 of the Appendix. Libraries were assessed for quality by high sensitivity DNA chip (Agilent, 5067-4626) on an Agilent 2100 bioanalyzer and quantified using Qubit dsDNA high sensitivity kit (Invitrogen, Q32851). Libraries were pooled manually and sequenced on an Illumina NextSeq 500 instrument using a NextSeq 500/550 High Output Kit v2.5 (150 cycles, illumina, 20024907) to produce 75bp paired end reads and 8bp dual index reads.

2.1.6 Bulk ATAC sequencing

Performed with assistance from Dr Christina Simoglou Karali

ATAC sequencing libraries were created following an in-house protocol based on the original published methods[173], [214]. 1000 monocytes or 1000 lineage negative $CD34^+$ HSPCs were sorted into transposition buffer, made up of digitonin, Tween-

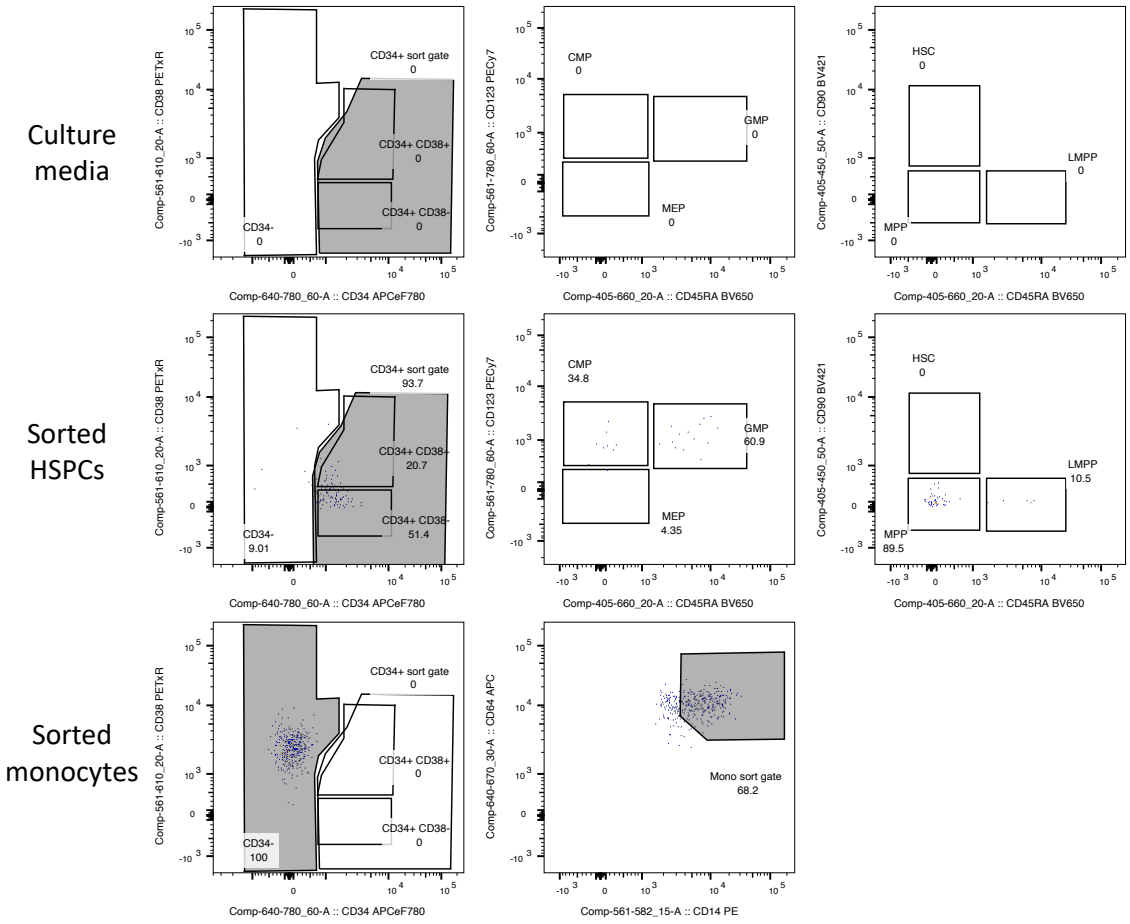


Figure 2.2: Assessment of sort purity - Cells, or culture media as a negative control, were sorted via the monocyte gate and HSPC gate respectively, collected and re-run to determine the accuracy of the initial sort

20, RNase-free PBS, UltraPure water and TD buffer (see Supplementary Table A.4 for volumes and suppliers). Once the sort had completed, TDE1 Tn5 transposase (illumina, 20034198) was added and plates were incubated at 37°C for 30 minutes. Transposase was removed using the MinElute Reaction Cleanup Kit (Qiagen, 28206) according to the manufacturer's recommendations and transposed DNA eluted in 12µl of UltraPure water (Invitrogen, 10977049). Libraries were indexed using custom HPLC-purified Nextera indexing primers (IDT) by amplification through 13-15 cycles of PCR using the indexing mix provided in Supplementary Table A.5. Libraries were cleaned up using AMPure XP beads (1.8:1 ratio) and quality assessed by high sensitivity DNA chip on a 2100 bioanalyzer. Libraries were quantified

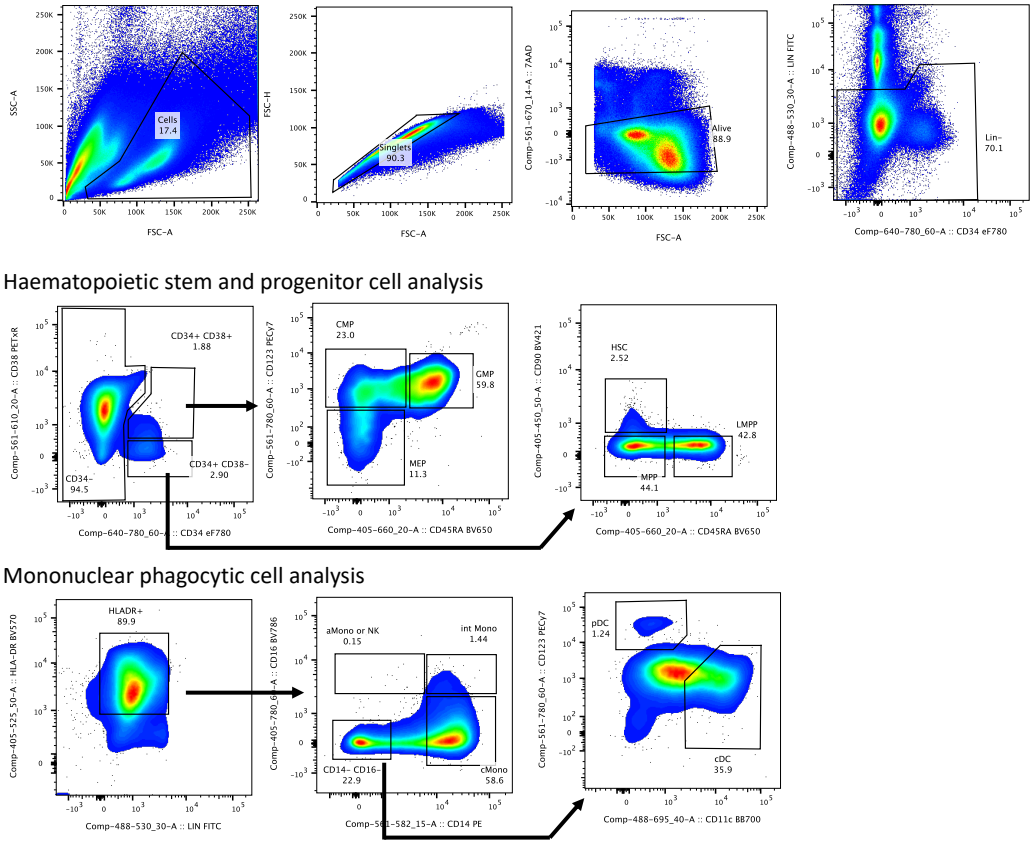


Figure 2.3: Gating strategy to identify HSPC subpopulations and mature myeloid populations - HSPC subpopulations - Haematopoietic stem cell (HSC, Lin$^-$ CD34$^+$ CD38$^-$ CD90$^+$ CD45RA$^-$), Multipotent progenitor cell (MPP, Lin$^-$ CD34$^+$ CD38$^-$ CD90$^-$ CD45RA$^-$), Lymphoid-primed multipotent progenitor (LMPP, Lin$^-$ CD34$^+$ CD38$^-$ CD90$^-$ CD45RA$^+$), Common myeloid progenitor (CMP, Lin$^-$ CD34$^+$ CD38$^+$ CD123$^+$ CD45RA$^-$), Granulocyte-monocyte progenitor (GMP, Lin$^-$ CD34$^+$ CD38$^+$ CD123$^+$ CD45RA$^+$), Megakaryocyte-erythroid progenitor (MEP, Lin$^-$ CD34$^+$ CD38$^+$ CD123$^-$ CD45RA$^-$), Classical monocyte (cMono, Lin$^-$ HLADR$^+$ CD14$^+$ CD16$^-$), intermediate monocyte (int Mono, Lin$^-$ HLADR$^+$ CD14$^+$ CD16$^+$), atypical monocyte (Lin$^-$ aMono, HLADR$^+$ CD14$^-$ CD16$^+$), classical dendritic cell (Lin$^-$ cDC, HLADR$^+$ CD14$^-$ CD16$^-$ CD123$^-$ CD11c$^+$), plasmacytoid dendritic cell (pDC, Lin$^-$ HLADR$^+$ CD14$^-$ CD16$^-$ CD123$^+$ CD11c$^-$)

using NEBNext Illumina library quantification kit (NEB, E7630S) and pooled for sequencing on an Illumina NextSeq 500 instrument using a NextSeq 500/550 High Output Kit v2.5 (75 cycles, illumina, 20024906) to produce 40bp paired end reads with an 8bp Read2 (i7) index read.

2.1.7 Single nuclear RNA and single cell ATAC sequencing

Performed with assistance from Dr Christina Simoglou Karali

Cryopreserved bone marrow mononuclear cells were thawed and stained with anti-CD34 as described above (Table 2.2). 10,000-25,000 alive, $CD34^+$ HSPCs or $CD34^-$ bone marrow MNCs were sorted into 50μl PBS with 0.2% BSA. Cells from three individuals were sorted into the same microcentrifuge tube to make up a maximum of 60,000 total cells. Samples were thawed and processed in three batches across two days to minimise batch effect. Nuclei were isolated in accordance with the low cell input nuclei isolation protocol (Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing Revision C, 10X Genomics, CG000365) with some modification of buffers to incorporate double the quantity of RNase inhibitor to each buffer (see Supplementary Tables A.6, A.7 and A.8). Sorted cells were centrifuged at 300G for 5 minutes at 4°C and supernatant removed. The cell pellet was re-suspended in ice cold lysis buffer and incubated for 2.5 minutes on ice for HSPCs and 3 minutes on ice for bone marrow MNCs. Lysis was inhibited by addition of wash buffer and nuclei separated by centrifuging at 500G for 5 minutes at 4°C. Supernatant was removed and nuclei were re-suspended in diluted nuclei buffer before being centrifuged again at 500G for 5 minutes at 4°C. Supernatant was removed completely and nuclei re-suspended in 5-6μl of diluted nuclei buffer. The entire nuclei suspension underwent transposition for 60 minutes at 37°C using reagents from the 10X Next GEM Single Cell Multiome ATAC kit A (10X Genomics, PN-1000280) and was loaded into Chip J (10X Genomics, PN-1000234) for gel bead-in emulsion (GEM) creation on the Chromium X controller. Single nuclear RNA-seq and single cell ATAC-seq libraries were created using the Chromium Next GEM Single Cell Multiome Reagent kit A (10X, PN-1000284) and

Table 2.2: Antibody panel for single nuclear RNA-seq and single cell ATAC-seq sort

Fluorophore	Target	Clone	Supplier	Titration (μl per 100μl)
APC-eFluor 780	CD34	4H11	Invitrogen	0.67
7AAD	Live/Dead		eBioscience (00-6993-50)	2

Library Construction Kit B (10X, PN-1000279) in accordance with the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide Revision F (CG000338). cDNA, indexed gene expression and indexed ATAC libraries underwent quality assessment using the high sensitivity DNA chip (Agilent, 5067-4626) on an Agilent 2100 bioanalyzer and were quantified using Qubit dsDNA HS kit (Invitrogen, Q32851) or NEBNext Illumina Library Quantification kit (NEB, E7630S). Libraries were sequenced by Novogene (www.novogene.com) on either a NovaSeq X Plus or NovoSeq 6000 instrument to obtain >20000 reads for gene expression libraries and >25000 reads for ATAC libraries per cell.

2.1.8 Antibody conjugation and optimisation for imaging mass cytometry

Unconjugated antibodies were conjugated using the MaxPar X8 antibody labelling kit (Standard Biotools, 201141A) in accordance with the manufacturer's instructions. Where possible, unconjugated antibodies that were free from bovine serum albumin (BSA) or glycerol were acquired to guarantee compatibility with the X8 labelling kit. This was not possible for anti-IL1RAP and therefore glycerol was removed using the BSA removal kit (Abcam, ab173231). The anti-IL1RAP antibody (Abcam, ab232950) was in a 56% glycerol solution and was therefore diluted 1 in 4 in UltraPure water and mixed with BSA removal buffer (0.8v/v). The entire solution was passed through a 50kDa filter in stages by centrifugation at 13,000G, which removed the glycerol (molecular weight 9Da) but retained the antibody. The antibody was then conjugated using the MaxPar X8 antibody labelling kit. All antibodies were titrated on tissue sections cut from the same blocks used for

the final acquisition. Images of stained sections were acquired using a Hyperion imaging mass cytometer (Standard BioTools), with all channels open to identify any spillover between channels. Antibody dilutions were selected to maximise signal to noise whilst minimising spillover into adjacent channels. Despite attempts to use primary anti-CSF1R and anti-CSF1 conjugated antibodies, signal remained very weak through multiple tested antibodies. Signal was therefore amplified using fluorophore-conjugated primary antibodies and metal-conjugated secondary antibodies. Isotype control non-targeting fluorophore-conjugated primary antibodies were tested alongside anti-CSF1R and anti-CSF1 antibodies and minimal background staining was detected. Emactuzumab was used as the anti-CSF1R antibody after first being conjugated with R-Phycoerythrin (R-PE) using the SiteClick antibody labeling kit (Invitrogen, S10467) in accordance with the manufacturer's protocol.

2.1.9 Imaging mass cytometry

4μm formalin fixed paraffin embedded bone marrow trephine sections were mounted on glass slides and baked at 60°C for 2 hours using a ThermoBrite (Leica) slide heater. All slides were processed and stained simultaneously to minimise batch effect. Slides were transferred immediately to two x 10 minute incubations in HistoClearII solution (National Diagnostics, HS-202). Tissue was rehydrated by four x 5 minute incubations in 100%, 100%, 95% and 70% ethanol. Slides were then washed twice in Milli-Q filtered water. Slides were transferred to pre-heated antigen retrieval solution (pH9, Dako, S2367) and incubated at 96°C using a steamer for 30 minutes. Slides were removed from the steamer and allowed to cool at room temperature for 10 minutes prior to transferring into Milli-Q water for two x five minute washes. Sections were encircled using a hydrophobic barrier pen and blocked using SuperBlock Blocking Buffer in PBS solution (Thermo Scientific, 37580) for 45 minutes at room temperature in a hydration chamber. Whilst slides were blocking an antibody master mix of fluorophore conjugated antibodies was prepared in MaxPar PBS (Standard Biotools, 201058), see Table 2.3) for full antibody details. Blocking buffer was tipped off and sections were stained with fluorophore conjugated

Table 2.3: Fluorophore conjugated antibody master mix for imaging mass cytometry

Fluorophore	Target	Clone	Supplier	Titration (μl per 100μl)
FITC	CSF1	D-4	Santa Cruz (sc-365779)	4
PE	CSF1R	Emactuzumab	SynOx Therapeutics	4

antibody master mix for 1 hour at room temperature. Slides were washed three times in MaxPar PBS and stained with metal conjugated antibody master mix (see Table 2.4 for full antibody details), including anti-fluorophore secondary antibodies, overnight at 4°C. The following day slides were washed in 0.1% Tween-20 in MaxPar PBS for 8 minutes, followed by two further five minute washes in MaxPar water (Standard Biotools, 201069). Excess water was tipped off slides and they were left in the dark until completely dry. Sections were imaged using a Hyperion imaging mass cytometer. One to three regions of interest were acquired from each section. Regions of interest were identified using paired haematoxylin and eosin stained images of each section and selected to be representative of the section as a whole. Prior to acquisition laser intensity was tested on an adjacent area to the region of interest to ensure optimal marker recovery.

2.2 Primary patient cell culture assays

2.2.1 Primary monocyte culture assays

Cryopreserved peripheral blood MNCs were thawed as outlined above, counted and centrifuged at 300G for 10 minutes at 4°C. Supernatant was removed and cells were resuspended in 100μl FACS buffer supplemented with 20μl of anti-CD3 microbeads (Miltenyi Biotec, 130-097-043) and 20μl of anti-CD19 microbeads (Miltenyi Biotec, 130-050-301). Cells were incubated at 4°C for 15 minutes then washed in 2ml FACS buffer and centrifuged at 300G for 10 minutes at 4°C. Monocytes were negatively enriched for by passing the cell suspension through an LS column (Miltenyi Biotec, 130-042-401) attached to a QuadroMACS separator and keeping the flow through. Monocyte enrichment was confirmed by flow cytometry during a pilot experiment. Negative enrichment was chosen to reduce the likelihood

Table 2.4: Metal conjugated antibody master mix for imaging mass cytometry

Metal	Target	Clone	Supplier	Titration (μl per 100μl)
115In	CD20	H1	Standard Biotools (91H013115)	2
141Pr	CD38	EPR4106	Standard Biotools (3141018D)	0.5
142Nd	CD56	123C3	Invitrogen (MA1-06801)	4
143Nd	HLADR	LN3	Invitrogen (14-9956-82)	0.25
144Nd	CD117	C117/370	Novus Bio (NBP2-29423)	4
145Nd	Fibronectin	Polyclonal	Novus Bio (NBP1-91258)	1.33
146Nd	CD16	EPR16784	Standard Biotools (3146020D)	1
147Sm	CD163	EDHu-1	Standard Biotools (3147021D)	0.25
148Sm	IL1B	Polyclonal	R&D systems (AF-201-NA)	8
149Sm	CD15	W6D3	Standard Biotools (3149026D)	1
150Sm	ITGA4	Polyclonal	Novus Bio (NB110-55525)	4
151Eu	VWF	VWF635	NeoBiotechnologies (7450-MSM2-P1ABX)	2
151Eu	CD31	EPR3094	Standard Biotools (91H027151)	0.67
152Sm	CD45	D9M8I	Standard Biotools (3152018D)	0.13
153Eu	FITC	FIT-22	BioLegend (408302)	0.67
154Eu	TGFBR1	RM0016-3A11	Novus Bio (NBP2-11921)	4
155Gd	Perilipin	Polyclonal	Novus Bio (NB110-40760)	0.33
156Gd	ITGB1	epr16895	Abcam (ab271931)	1
158Gd	CEBPB	21A1	Novus Bio (NBP1-41216)	4
159Tb	CD68	KP1	Standard Biotools (3159035D)	0.08
160Dy	CD34	QBend/10	Thermo Scientific (MA1-10202)	1.33
161Dy	HLF	M01	Abnova (H00003131-M03)	0.5
162Dy	IL1R1	Polyclonal	Abcam (ab190078)	0.5
163Dy	CD14	SP192	Abcam (ab230903)	0.25
164Dy	TGFb	Polyclonal	Novus Bio (NBP1-80289)	1.67
165Ho	FLT3	7B7C3	GeneTex (GTX60777)	2
166Er	CD45RA	HI100	Standard Biotools (3166031D)	0.33
167Er	MET	D1C2	Standard Biotools (3167020D)	1
168Er	Ki-67	B56	Standard Biotools (3168007B)	0.25
169Tm	IL1RAP	Polyclonal	Abcam (ab232950)	2
170Er	CD3	Polyclonal	Standard Biotools (3170019D)	0.63
171Yb	PE	PE001	BioLegend (408102)	1
172Yb	HGF	7-2	Novus Bio (NBP1-19182)	4
173Yb	CD271	EP1039Y	Abcam (ab256584)	2
175Lu	CD235ab	HIR2	Standard Biotools (3175029D)	0.08
176Yb	CD123	6H6	Abcam (ab21562)	8
195Pt	ICSK1	Segmentation mix	Standard Biotools (201500)	0.5
196Pt	ICSK2	Segmentation mix	Standard Biotools (201500)	0.5
209Bi	aSMA	1A4	Standard Biotools (3141017D)	0.08

of monocytes being altered by the effects of antibodies binding to their surface. Cells were counted and centrifuged at 400G for 5 minutes at room temperature. Supernatant was removed and cells were resuspended at 1-2x10^6 cells/ml in culture media (RPMI1640 supplemented with 10% FBS, 1% penicillin/streptomycin and L-glutamine). Cells were seeded into 6 well plates and rested for 3-6 hours at 37°C in a humidified incubator containing 5% CO_2. Cells were removed from culture and transferred to 50ml cylindrical tubes. Cells were pelleted by centrifugation at 300G for 10 minutes and supernatant removed. Cells were resuspended in culture medium at 2x10^5/ml and 10000 cells per well transferred to pre-prepared 96 well culture plates. Culture media within the plates was supplemented with a range of doses of anti-CSF1R humanised monoclonal antibody emactuzumab (SynOx therapeutics). Some wells contained 100ng/ml MCSF (Peprotech, 300-

Table 2.5: Antibody panel for monocyte immunophenotyping

Fluorophore	Target	Clone	Supplier	Titration (μl per 100μl)
BV711	CD115	9-4D2-1E4	BioLegend (347318)	2
BV510	CD14	63D3	BioLegend (367123)	1
APC	CD16	B73.1	BioLegend (360705)	0.31
FITC	CD3	SK7	BD (345763)	2
FITC	CD19	HIB19	eBioscience (11-0199-42)	2
Zombie Red	Live/Dead		BioLegend (423109)	5
PE	CSF1R	Emactuzumab	SynOx Therapeutics	2
PE-Cy7	CD56	5.1H11	BioLegend (362509)	0.63

25) and others contained 100ng/ml of GM-CSF (Peprotech, 300-03). Culture plates were returned to the incubator for up to 72 hours and assayed using either CellTiter-Glo 2.0 (Promega, G9242) or CellTox Green (Promega, G8741) to detect viable/apoptotic cell numbers within each well by luminescence or fluorescence detection respectively on a CLARIOStar Plus microplate reader (BMG Labtech). Excess cells not required for culture assays were assayed by flow cytometry on an Attune NxT instrument using the monocyte immunophenotyping antibody panel (Table 2.5) using the staining protocol described above.

2.2.2 Primary HSPC proliferation assay

Cryopreserved bone marrow mononuclear cells were thawed and counted as previously described. Cells were centrifuged at 300G for 10 minutes, supernatant removed and resuspended at 2x10^6 cells/ml in recovery media (StemSpan SFEM (StemCell Technologies, 09650), 20% BIT9500 (StemCell Technologies, 9500), 1% penicillin/streptomycin, 50ng/ml FLT3L (Peprotech, 300-19), 50ng/ml SCF (Peprotech, 300-07), 10ng/ml TPO (Peprotech, 300-18-10) and 10ng/ml IL-3 (Peprotech, 200-03). Cells were cultured overnight at 37^oC in a humidified incubator containing 5% CO_2. The following day they were centrifuged at 300G for 10 minutes, supernatant was replaced with 100μl FACS buffer supplemented with 5% TruStain FcX (BioLegend, 422302) and incubated on ice for 10 minutes. Cells were stained by adding the antibody cocktail in Table 2.6. 500 lineage negative $CD34^+$ HSPCs

Table 2.6: Antibody panel for HSPC culture sort

Fluorophore	Target	Clone	Supplier	Titration (μl per 100μl)
BV421	CD90	5E10	BioLegend (328122)	3.33
BV650	CD45RA	HI100	BioLegend (304136)	2
FITC	Lineage cocktail	RPA-2.10, OKT3, 61D3, CB16, HIB19, TULY56, HIR2	eBioscience (22-7778-72)	5
PE-Texas Red	CD38	HIT2	Invitrogen (2208339)	5
PE/Cy7	CD123	6H6	eBioscience (25-1239-42)	1.67
APC-eFluor 780	CD34	4H11	eBioscience (47-0349-42)	1

were sorted, using a FACSAria Fusion instrument, into each well of a 96 well round bottomed culture plate. Culture plates were prepared in advance with 100μl/well of base media supplemented with specific cytokine and inhibitor combinations for each condition being tested. These conditions are outlined in Table 2.7. Cells were cultured for 8 days, with cytokines being replenished every 72 hours. 50μl of media was removed from each well and antibody cocktail added. Cells were stained in the dark for 30 minutes at 4^oC , washed in FACS buffer and then centrifuged at 400G for 5 minutes. Supernatant was removed and cells were resuspended in 200μl of FACS buffer supplemented with 7AAD and CountBright plus Absolute Counting Beads (Invitrogen, C36995). Plates were assayed by flow cytometry using an Attune NxT instrument (Thermo Scientific). Cell count was calculated by normalising total alive cell number to counting beads. Response to each cytokine/inhibitor was determined by comparison between conditions after first normalising cell count to the untreated control wells for each individual. Differentiated cell populations were identified by immunophenotyping as shown in Figure 2.4.

2.2.3 Primary HSPC survival assay

Cryopreserved bone marrow mononuclear cells were thawed and cultured overnight in recovery media as described above. 500 lineage negative $CD34^+$ HSPCs were

Table 2.7: Culture media for primary HSPC proliferation assays

Base media for all assays		
Component	**Concentration**	**Supplier**
StemSpan SFEM		StemCell Technologies (09650)
BIT9500	20%	StemCell Technologies (9500)
Penicillin/Streptomycin	1%	Gibco (15140122)
MCSF assay		
FLT3L	10ng/ml	Peprotech (300-19)
SCF	10ng/ml	Peprotech (300-07)
IL-3	10ng/ml	Peprotech (200-03)
M-CSF OR 0.1% BSA	100ng/ml	Peprotech (300-25)
Emactuzumab OR Vehicle	100ng/ml	SynOx Therapeutics
IL1β assay		
FLT3L	10ng/ml	Peprotech (300-19)
SCF	10ng/ml	Peprotech (300-07)
IL-3	10ng/ml	Peprotech (200-03)
IL-1β	0.1, 1 or 10ng/ml	Peprotech (200-01B)
TGFβ assay		
FLT3L	50ng/ml	Peprotech (300-19)
SCF	50ng/ml	Peprotech (300-07)
TPO	10ng/ml	Peprotech (300-18)
IL-3	10ng/ml	Peprotech (200-03)
IL-6	10ng/ml	Peprotech (200-06)
G-CSF	10ng/ml	Peprotech (300-23)
GM-CSF	5ng/ml	Peprotech (300-03)
EPO	1iu/ml (6.25ng/ml)	Peprotech (160-64)
TGFβ	0.1, 1 or 10ng/ml	Peprotech (100-21)

Table 2.8: Antibody panel for HSPC culture output analysis

Fluorophore	Target	Clone	Supplier	Titration (μl per 1ml)
BV421	CD235a	GA-R2	BD Biosciences (562938)	2
FITC	CD15	MMA	BioLegend (394706)	2
PE	CD14	M5E2	BioLegend (301806)	3.33
PE-Cy7	CD33	P67.6	BioLegend (366618)	2.5
AF700	CD71	M-A712	BD Pharminogen (563769)	2.5
APC-eFluor780	CD34	4H11	Invitrogen (47-0349-42)	1

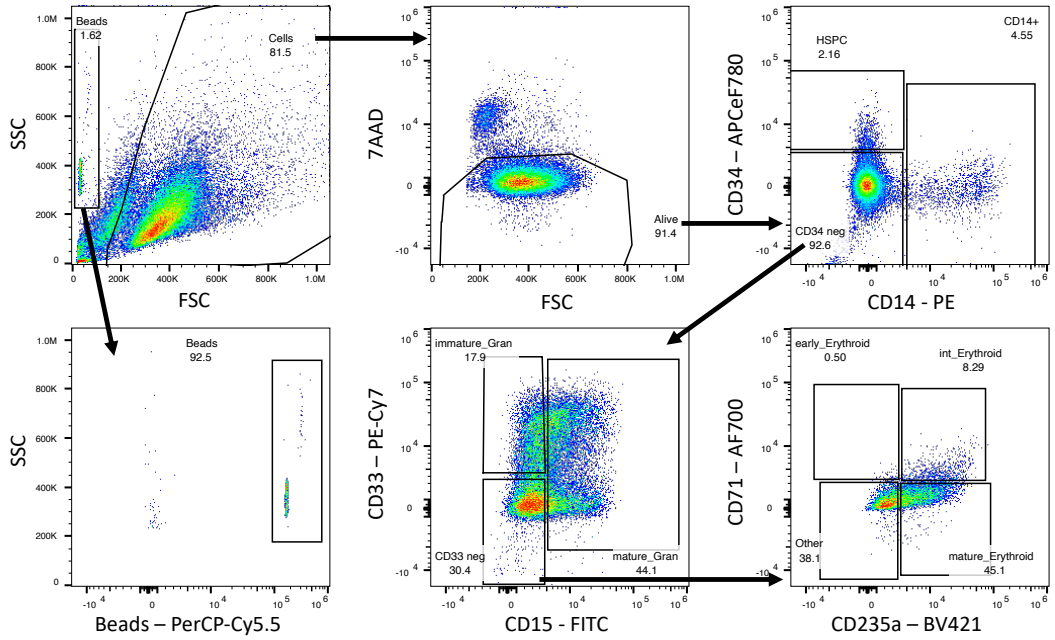


Figure 2.4: Liquid culture output assessment for HSPC primary culture assays - 500 lineage negative CD34 positive cells were cultured per well for 8 days (proliferation) and 13 days (survival) in serum free conditions. Representative plots demonstrate gating for output analysis. Alive cells were normalised using counting beads, which were gated from the PerCP-Cy5.5 channel. Differentiation output quantified by identifying $CD34^+$ HSPCs, $CD14^+$ monocytes, $CD33^+$ $CD15^-$ Immature granulocytes, $CD15^+$ $\pm$CD33 Mature granulocytes, $CD33^-$ $CD71^+$ $\pm$CD235a Immature erythroid, $CD33^-$ $CD71^-$ $CD235a^+$ Mature erythroid

sorted, using a FACSAria Fusion instrument, into each well of a 96 well round bottomed culture plate. Culture plates were prepared in advance with 100μl/well of antibiotic free base media (StemSpan SFEM with 20% BIT9500) without additional supportive cytokines other than the specific cytokine and inhibitor combinations for each condition being tested. These conditions are outlined in Table 2.9. Cells were cultured in these minimal cytokine conditions for 48 hours and then a highly proliferative cytokine cocktail was added (see Table 2.9). Cells were cultured for a further 11 days with cytokines being replenished every 72 hours. Cells were then stained and assayed as previously described in the primary HSPC proliferation assay.

Table 2.9: Culture media for primary HSPC survival assays

Base media for all assays		
Component	**Concentration**	**Supplier**
StemSpan SFEM		StemCell Technologies (09650)
BIT9500	20%	StemCell Technologies (9500)
MCSF assay		
M-CSF OR BSA	100ng/ml	Peprotech (300-25)
Emactuzumab OR Vehicle	0.1 or 10μg/ml	SynOx Therapeutics
TGFβ assay		
FLT3L	10ng/ml	Peprotech (300-19)
SCF	10ng/ml	Peprotech (300-07)
TGFβ OR 0.1% BSA	10ng/ml	Peprotech (100-21)
Proliferative cytokine cocktail - added to all assays after 48hr		
Penicillin/Streptomycin	1%	Gibco (15140122)
FLT3L	50ng/ml	Peprotech (300-19)
SCF	50ng/ml	Peprotech (300-07)
TPO	10ng/ml	Peprotech (300-18)
IL-3	10ng/ml	Peprotech (200-03)
IL-6	10ng/ml	Peprotech (200-06)
G-CSF	10ng/ml	Peprotech (300-23)
GM-CSF	5ng/ml	Peprotech (300-03)
EPO	1iu/ml (6.25ng/ml)	Peprotech (160-64)

2.3 Bioinformatic analysis

2.3.1 Bulk ATAC-seq analysis

Preprocessing pipeline was written and run by Sean Wenn

Fastq files were trimmed using trimgalore! v0.6.5 to remove poor quality bases and residual illumina adapter sequences[215]. Bowtie2 v2.4.2 was used to align reads to GENCODE v38 primary assembly reference genome[216]. Uniquely mapped reads were isolated using samtools v1.9 and PCR duplicates removed using picard-tools v2.3.0[217], [218]. Mitochondrial reads were also removed using samtools. ATAC-seqQC v1.14.4 was used to shift coordinates at Tn5 insertion sites and create plots for performing quality control assessment[219]. Quality of ATAC-seq libraries was assessed by visualising fragment length distribution graphs, heatmaps of nucleosome

free and mononucleosome fragments mapped to regions surrounding transcription start sites and CTCF transcription factor footprints within mapped fragments. Examples of these plots are shown in Figure2.5. Libraries were also compared for sequencing depth, mapping rates, mitochondrial reads and PCR duplicate reads. All sequenced libraries were found to be of sufficient quality to be included in the analysis. MACS2 v2.2.7.1 was used to call peaks using –shift -75 –extsize 150 –nomodel –call-summits –nolambda –keep-dup all -p 0.01 settings[220]. Peaks were collapsed initially by sample, to create a non-overlapping peak set, and then across samples, to create a consensus peak set, using R v4.0.1 and the GenomicRanges package. Peaks were filtered to remove those overlapping ENCODE blacklist regions, those that were overhanging the start or end of chromosomes and any non-reproducible peaks, where fragments were only identified from a single sample. This peak by count matrix was exported into R v4.3.1 and annotated using ChIPseeker v1.36.0[221]. Differential peak expression was performed using DESeq2 v1.40.2[222] and principal component analysis (PCA) by PCAtools v2.12.0(https://github.com/kevinblighe/PCAtools). PCA plots were used to assess for batch effect between sequencing runs and sort dates, as shown in Figure 2.6. No significant batch effect was found and therefore batch was not included in the design matrix for DESeq analysis. Transcription factor motif enrichment analysis was performed using chromVAR v1.22.1[223]. Position weighted matrices for transcription factor binding motifs used in this analysis were taken from JASPAR2022 v0.99.7 package[224]. Visualisations were produced using corrplot v0.92 (https://github.com/taiyun/corrplot), ggplot2 v3.4.4[225] and ComplexHeatmap v2.16[226].

2.3.2 SmartSeq3 bulk RNA-seq analysis

Fastq files were created from bcl2 sequencing files using bcl2fastq v2.20.0.422. Unique Molecular Identifier (UMI) containing reads were identified, and UMI sequences extracted, using umi-tools v1.1.2[227]. Regular expression equations were used with umi-tools extract to identify 8bp UMI sequences following GCGCAATG on either Read1 or Read2:

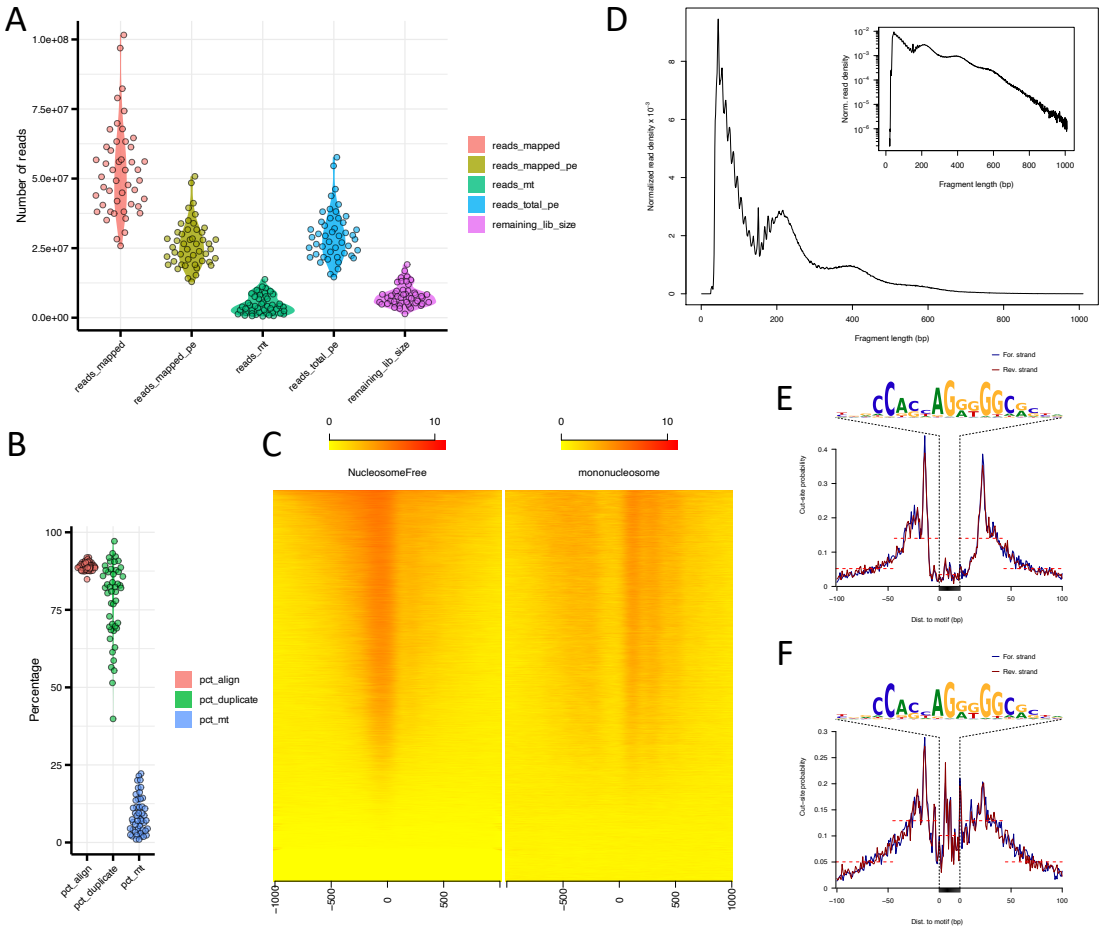


Figure 2.5: ATAC-seq quality control assessment - Violin plots of sequencing metrics by reads (A) and percentages (B). reads_mapped - number of reads successfully mapped to genome, reads_mapped_pe - number of paired end reads mapped successfully to genome, reads_mt - reads mapped to mitochondrial genes, reads_total_pe - total paired end reads, remaining_lib_size - Number of non-duplicate non-mitochondrial reads available for analysis, pct_align - percentage aligned, pct_duplicate - percentage of aligned reads that are PCR duplicates, pct_mt - percentage mitochondrial reads. (C) Representative heatmap of nucleosome free and mononucleosome fragment distribution around transcription start sites. (D) Fragment length distribution demonstrating nucleosome free reads (<100bp) and periodic nucleosome containing peaks every 200bp. High quality CTCF footprint (E) and lower quality CTCF footprint (F)

```
--bc-pattern = "\^(?P<discard_1>.*)(?P<discard_2>GCGCAATG)
      {s<=1}(?P<umi_1>.{8})(?P<discard_3>GG.).*"
--bc-pattern2 = ".*(?P<discard_4>.GG)(?P<umi_2>.{8})
      (?P<discard_5>GTAACGCG){s<=1}(?P<discard_6>.*)$"
```

UMI containing reads were trimmed for poor quality bases and adapter sequences

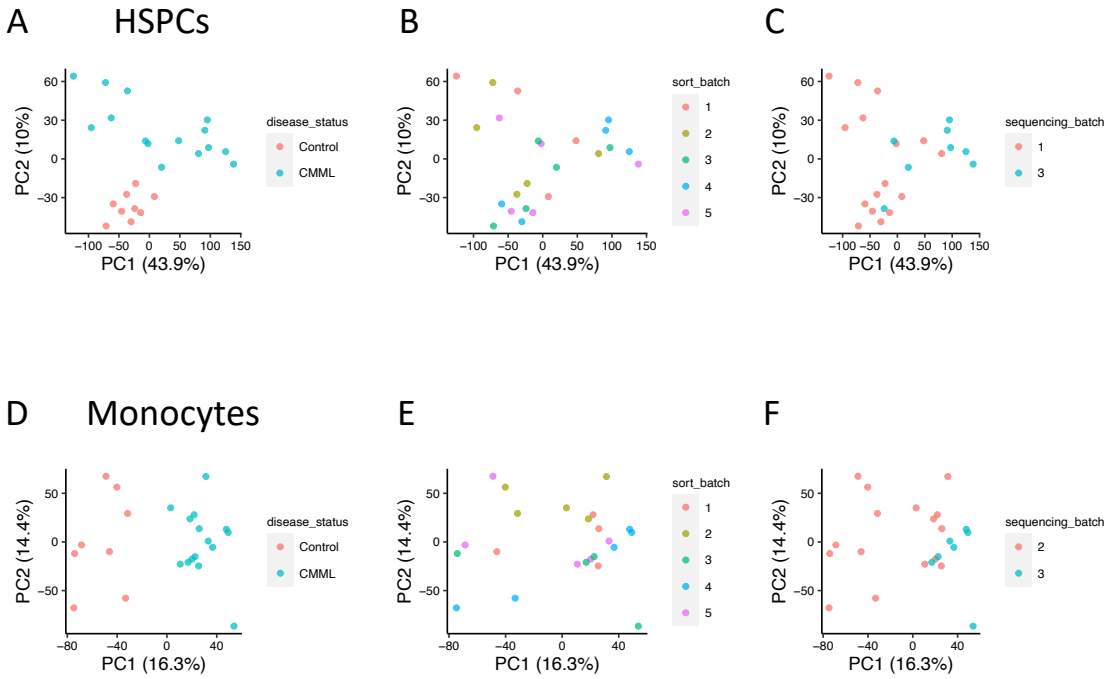


Figure 2.6: ATAC-seq batch effect assessment - Principal component analysis of ATAC-seq samples from Lin$^-$ CD34$^+$ HSPCs (A-C) and CD14$^+$ CD64$^+$ monocytes (D-F) comparing samples between TET2m CMML patients (CMML) and healthy donors (Control) (A and D), sort date (B and E) and sequencing pool (C and F).

using trimgalore! v0.6.5[215]. Reads were aligned to GENCODE v38 primary assembly genome with STAR v2.7.3a using default settings with –twopassMode basic and BAM files sorted by coordinate [228]. BAM files were indexed using samtools and then umi-tools dedup was used to remove PCR duplicates based on co-ordinates. Subread's featureCounts package was used to count reads aligned to gene meta-features using default parameters[229]. Sequencing quality control was performed using fastqc v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and results synthesised using multiqc v1.13[230]. Count matrices were passed into R v4.3.1 and a further round of quality control was performed. Libraries with fewer than 6500 expressed genes were excluded from further analysis (Figure 2.7A and C). Visual inspection of PCA analysis of TMM normalised log2 CPM counts, shown in Figure 2.7B and D, found evidence of batch effect for sort date in both HSPC and monocyte samples. Batch correction was performed using limma to

account for library size, mitochondrial reads and sort batch[231]. Differential gene expression analysis was performed using DESeq2 with mitochondrial reads and sort batch included in the design formula[222]. Pathway analysis was performed using ClusterProfiler v.4.8.2[232] for over representation analysis and fast GSEA (fgsea) v.1.26.0 for gene set enrichment analysis[233]. Differentially expressed genes (alpha = 0.1) were used as input for over representation analysis. Gene set enrichment analysis was performed on a ranked gene list, ranked by DESeq2 likelihood ratio statistic comparing populations of interest. Differential expression analysis between TET2m CMML and healthy donors was performed with an alpha of 0.05. The more stringent p value threshold was chosen because of the larger sample size for this analysis.

2.3.3 Survival analysis

Data was collected on survival or most recent follow up in all patients assayed by bulk ATAC-seq. Kaplan-Meier analysis comparing survival between chromatin accessibility defined, phenotypic clusters was performed using the survminer v0.4.9 package (https://github.com/kassambara/survminer). Multivariable Cox proportional hazard regression analysis was performed using the coxph function of the survival v3.5-7 package[234].

2.3.4 Single nucleotide polymorphism identification for demultiplexing individuals from 10X Multiome samples

Fastq files from bulk RNA-seq of both monocyte and HSPC populations were concatenated to create a single fastq file for each individual. Poor quality bases, sequencing adapters and UMI barcodes were removed using cutadapt v3.1 with options -j 2 -a "CTGTCTCTTATACACATCT;o=5" -g "ATTGCGCAATGNNNNNNNNGGG;o=8" -a "A50;o=20" -m 10[235]. Reads were aligned to the GENCODE v38 GRCh38 primary assembly genome using STAR v2.7.3a with –sjdbOverhang 74 and –twopassMode Basic[228]. PCR duplicates were labelled on coordinates by picard-tools v2.3.0. Single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) were identified using GATK 4.3.0.0 in accordance with the GATK best practice

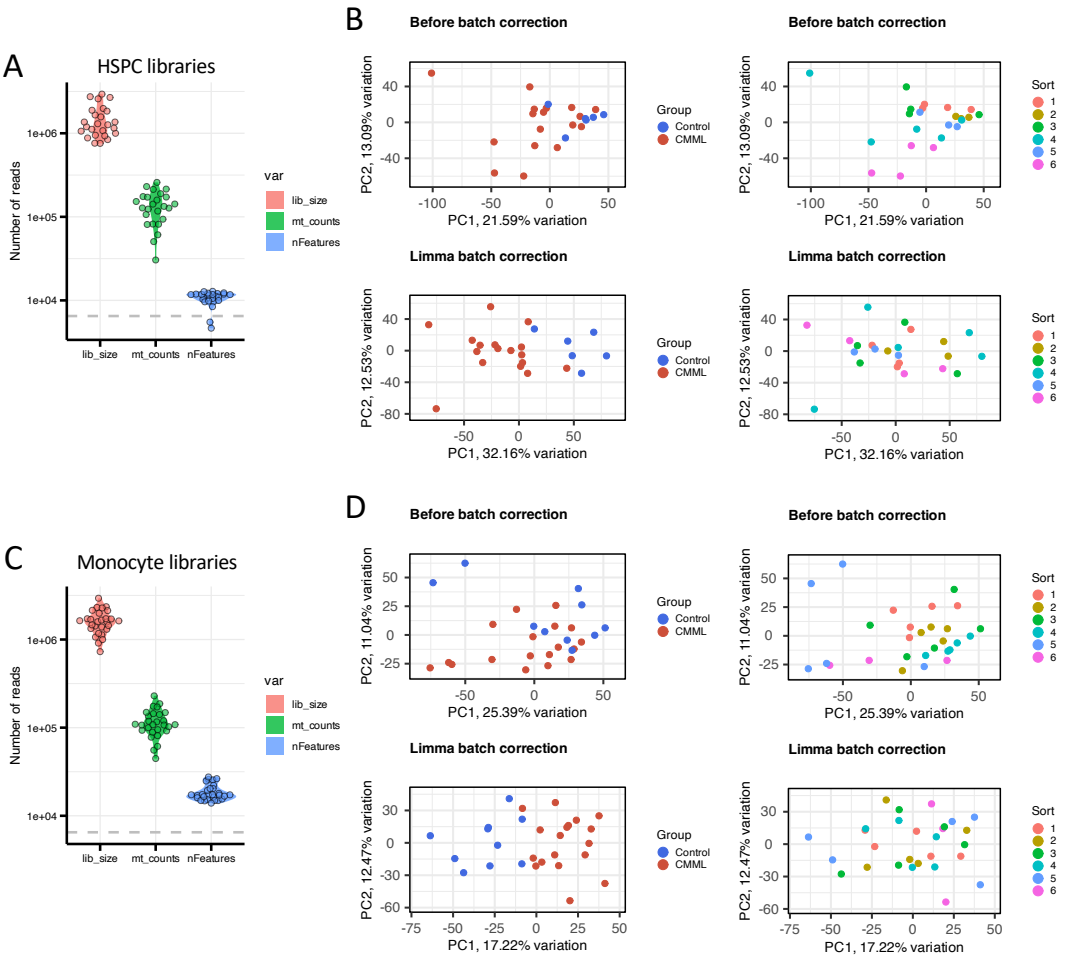


Figure 2.7: RNA-seq quality control and batch effect assessment - Quality control metrics for RNA-seq libraries prepared from RNA extracted from sorted Lin$^-$ CD34$^+$ HSPCs (A) and CD14$^+$ CD64$^+$ monocytes (C). lib_size - reads per sample, mt_counts - counts mapped to mitochondrial genes and nFeatures - number of unique expressed genes. Libraries with fewer than 6500 expressed genes (nFeatures) were excluded from further analysis. PCA plots used for visual assessment of batch effect from sort date in HSPC (B) and monocytes (D) libraries. Top panels show plots for PCA calculated on uncorrected normalised data, bottom panels show plots for PCA calculated on limma corrected counts corrected for library size, mitochondrial reads percentage and sort batch. Left panels coloured by disease group, right panels coloured by sort batch.

pipeline for RNA-seq calling of germline genetic variants (https://github.com/gatk-workflows/gatk4-RNA-seq-germline-snps-indels). This involved running the AddOrReplaceReadGroups, SplitNCigarReads, BaseRecalibrator, ApplyBQSR and HaplotypeCaller tools using default settings. Output variant call format (VCF) files were merged using bcftools v1.17 and filtered using GATK v4.3.0.0 VariantFiltration command with –window 35 –cluster 3 –filter-name "FS" –filter-expression "FS > 30.0" –filter-name "QD" –filter-expression "QD < 2.0" settings.

2.3.5 Single nuclear RNA-seq and single cell ATAC-seq analysis

Alignment, ambient RNA removal, genetic demultiplexing of individuals and doublet detection

Fastq files were aligned to the 10X curated cellranger-arc GRCh38-2020-A-2.0.0 reference genome and features were counted using cellranger-arc v2.0.2 count command with default settings. Gene expression (GEX) matrices were extracted using a custom script based on scripts used in the Nextflow nf-cellbender pipeline (https://github.com/cellgeni/nf-cellbender) and CellBender v0.3.0 applied. CellBender creates a model of ambient RNA expression for each sample using machine learning, with which it then identifies and removes ambient RNA counts from GEX matrices and filters out empty/ambient droplets[236]. This is particularly important for snRNA-seq because nuclei isolation releases large amounts of cytoplasmic RNA from lysed cells, which can stick to nuclei and become incorporated into the GEM, creating ambient RNA contamination. Genetic demultiplexing and doublet detection was performed using the Demuxafy pipeline[237]. Souporcell was used to demultiplex individuals using the –common_variants option[238]. VCF files created by Souporcell were correlated to VCF files created using the GATK best practice pipeline for RNA-seq calling of germline genetic variants (described above) using Demuxafy Assign_Indiv_by_Geno.R script. Souporcell was also used to identify heterotypic doublets, droplets containing multiple cells from different individuals, and to predict the global percentage of ambient RNA for each sample. Droplet doublets, containing two or more cells from the same individual, were identified

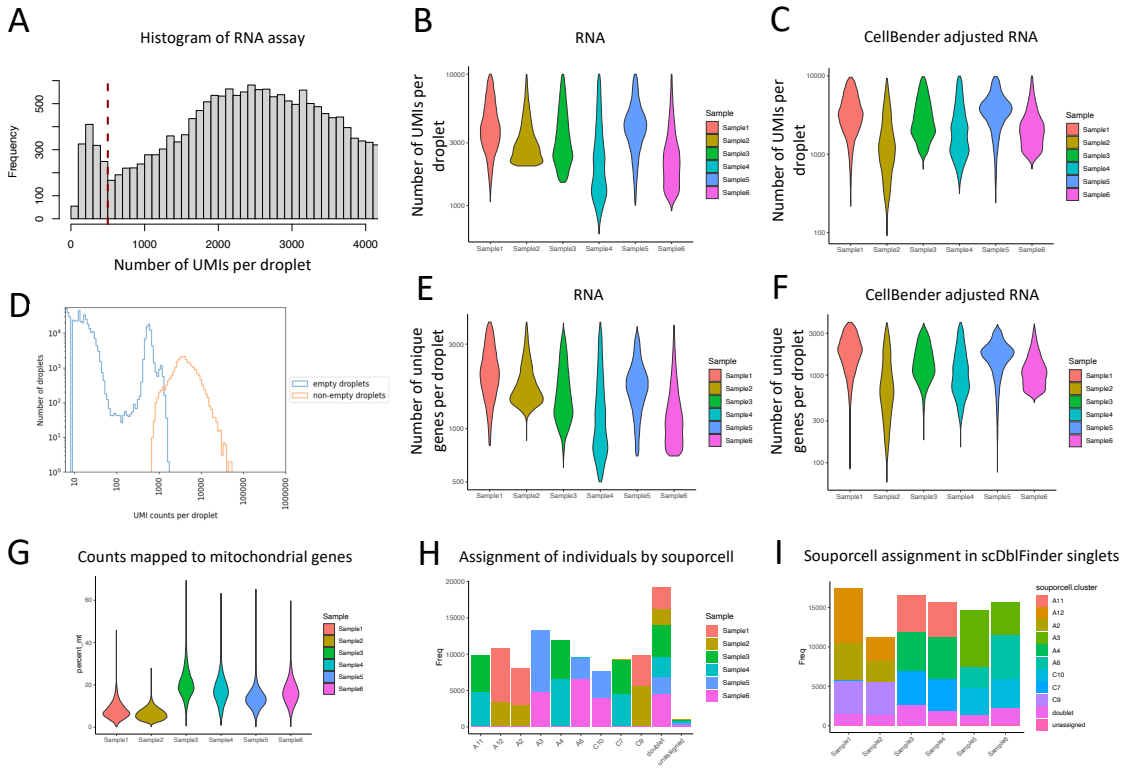


Figure 2.8: snRNA-seq quality control assessment - (A) Representative histogram of UMI counts per droplet with manual threshold shown by red vertical dotted line. Violin plots showing a comparison of RNA (B and E) and CellBender adjusted RNA (C and F) UMI counts and expressed genes by sample. (D) CellBender output showing UMI count by droplet. Empty and non-empty droplets are displayed in blue and orange respectively. (G) Percentage of reads mapping mitochondrial genes plotted by sample. (H) Souporcell assigned droplets coloured by sample, split by individuals, heterotypic doublets and unassigned droplets. (I) Stacked bar plot coloured by individual, scDblFinder doublets and scDblFinder unassigned droplets showing number of cells assigned to each group split by sample.

by running scDblFinder and scrublet [239], [240]. scDblFinder was also run on scATAC-seq data using the aggregateFeatures=TRUE parameter as recommended by the authors. Souporcell individual assignments and doublet detection results were combined using the Demuxafy Combine_Results.R script.

Quality control and the removal of poor quality droplets

I performed an initial exploratory analysis on each sample prior to merging samples together. Each sample, which corresponds to a single lane run on the Chromium X controller, had pre-processed count data passed into Seurat v4.4.0 for quality

control analysis[241]. As explained above, CellBender removes empty or ambient droplets based on RNA expression, I therefore used the list of barcodes that passed CellBender filtration as the backbone of the Seurat object to which I added unfiltered cellranger-arc RNA counts and extracted data on number of UMI reads, number of expressed genes and reads mapping to mitochondrial DNA (see Figure 2.8). ATAC data was similarly passed into Signac v1.10.0 and data was extracted on number of UMI counts, number of mapped peaks, nucleosome signal and transcription start site score (see Figure 2.9)[242]. Plots were created from these data, which I manually inspected to determine sample specific thresholds for filtering low quality droplets. I decided against applying a universal lower threshold for all samples because empty/ambient droplets were removed by CellBender, and because there was variability between samples and batches that were secondary to unequal sequencing depth or cell type composition. Universal thresholds were however applied to remove droplets with over 10000 UMI counts or expressed genes as these were highly likely to be doublets. Droplets were also filtered on ATAC-seq quality control parameters, removing droplets with TSS enrichment below 3 or nucleosome signal above 2. Doublets were removed if detected by souporcell or both scDblFinder and scrublet on RNA analysis. All doublets were removed if detected by scDblFinder on ATAC analysis.

Each sample was automatically annotated for cell type using SingleR and differences between raw RNA counts and CellBender adjusted RNA counts were investigated by plotting cell type marker genes on UMAP projections and cell type assignment to clusters for each sample[243]. Visual inspection of these plots found that monocytes and normoblasts were responsible for most of the ambient RNA, and CellBender successfully removed much of the noise this created and improved automated cell type annotation for all samples. Representative plots are shown in Figure 2.10. CellBender adjusted RNA was therefore used for downstream analysis.

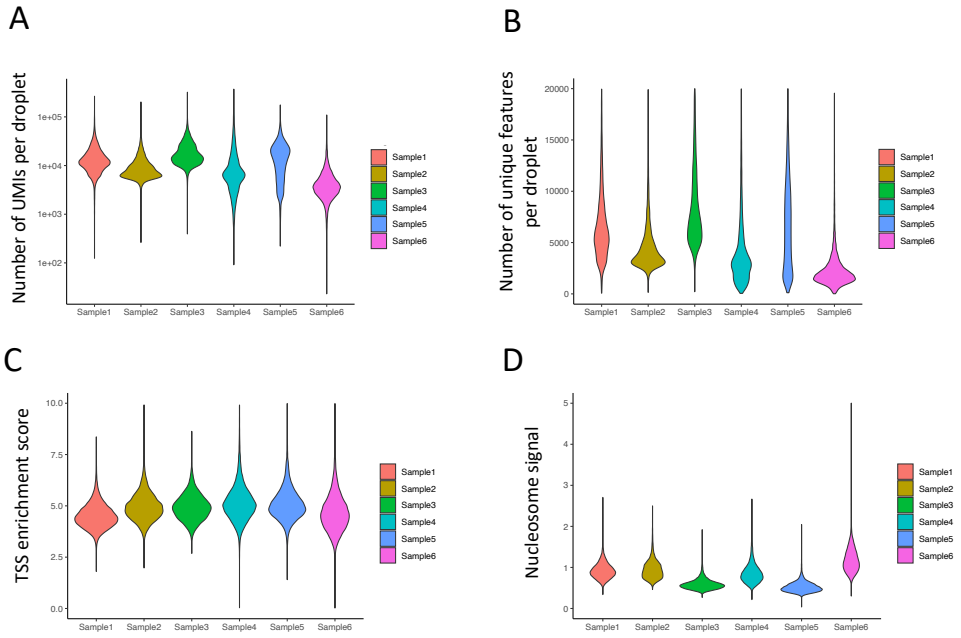


Figure 2.9: scATAC-seq quality control assessment - Violin plots by sample for UMI counts per droplet (A), number of fragments aligning to unique loci (B), transcription start site enrichment score (C) and nucleosome signal (D).

Merging samples and data integration

Samples were merged to create a single Seurat object. As described above, RNA and CellBender adjusted RNA count matrices were combined for each individual sample and filtered. A consensus peak set was created using the GenomicRanges v1.52.0 package by merging BED files from each of the samples containing peak coordinates called by cellranger-arc. ATAC-seq fragment data were combined with the consensus peak set to create a feature matrix using Signac, and added to the Seurat objects containing RNA data for the same cells. Seurat objects were then merged to create a single combined object and MACS2 was used to call peaks on the ATAC data. Peaks that overlapped coordinates outside of standard chromosomes and ENCODE blacklist regions were filtered and the peak-count matrix was added to the combined Seurat object.

Exploratory analysis of the combined object found significant variability between batches and individuals. Cellbender adjusted RNA counts were log normalised,

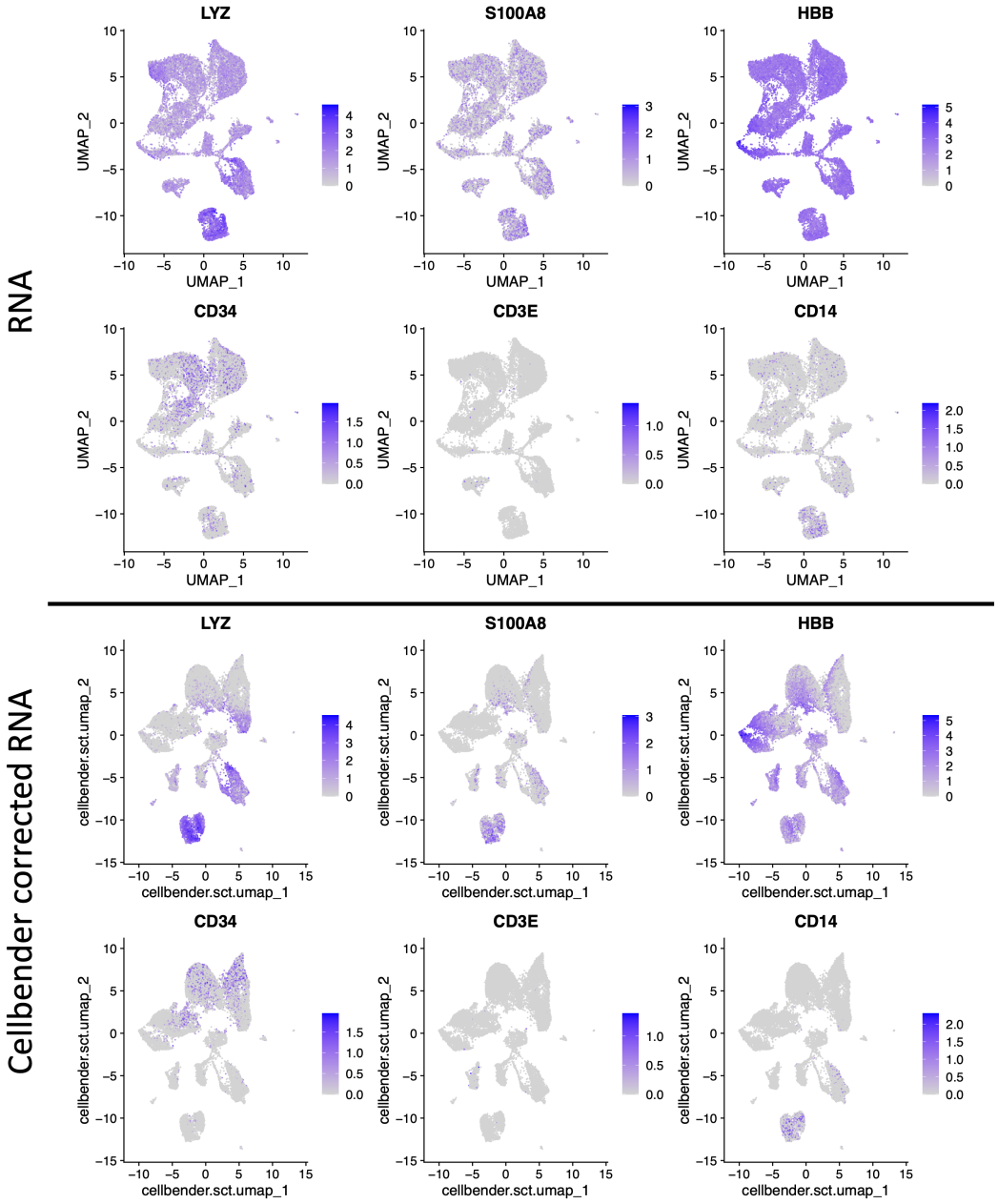


Figure 2.10: Comparison of marker gene expression using RNA or CellBender adjusted RNA counts - Log normalised RNA counts (top) and Log normalised CellBender adjusted RNA counts are plotted on UMAP plots for a representative sample. CD14, Lysozyme (LYZ) and S100A8 are markers for monocytes, HBB for normoblasts, CD34 for HSPCs and CD3E for T cells.

the top 2000 most variable features were identified, counts were scaled and PCA performed using default settings as part of a standard Seurat workflow. The count matrix for fragments in MACS2 called peaks was used for ATAC analysis. Counts were transformed for further analysis by latent semantic indexing (LSI)[244]. They were first normalised using term frequency-inverse document frequency on peaks present in more than 5 cells and then dimensionality reduction was performed using singular value decomposition by Signac. PCA and LSI matrices were used as input for Harmony v1.0.3 to remove variance from inter-individual differences, so that cells of similar cell type were brought together in reduced dimension space for annotation[245]. Weighted nearest neighbour (WNN) integration was then implemented to combine harmony adjusted PCA and LSI matrices to create a combined multimodal WNN dimensionality reduction matrix. UMAP projection, for visualisation, and unsupervised louvain clustering were performed on this matrix to identify multimodal defined cell states. Multiple different clustering resolutions were tested and the clustree v0.5.0 package was used to help identify an optimal resolution for annotation.

Cell type annotation and ambient RNA cluster removal

Automated annotation was performed using SingleR to reference data sets from the celldex package [175], [243], [246] and by mapping onto the Azimuth bone marrow reference dataset using Seurat[247]–[250]. Average RNA expression from each cluster was extracted and uploaded to CELLiD prediction using DISCO[249]. Over representation analysis was performed on haematopoietic cell gene sets curated within the SingleCellR package[251] and cell cycle gene set enrichment analysis was performed using the CellCycleScoring command in Seurat, to identify cells within different phases of the cell cycle. Data from each of these tools were manually interpreted and checked by plotting cannonical cell type marker genes before labelling clusters. Clusters that were highly similar were merged together to make both coarse and fine annotation labels. Annotation was performed using RNA gene expression data, rather than gene scores calculated from chromatin

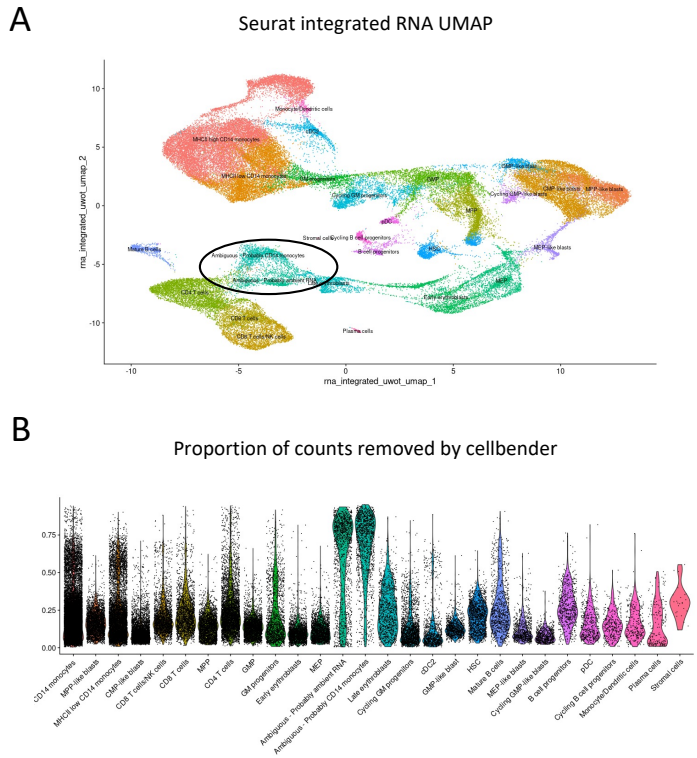


Figure 2.11: Discovery of ambiguous clusters containing cells with high levels of ambient RNA contamination - (A) UMAP projection of cellbender adjusted RNA counts integrated using Seurat to remove batch effect. Clusters of ambiguous cell type are highlighted in the black ring. (B) Violin plot of each cluster showing the proportion of counts removed by CellBender, a surrogate marker of ambient RNA contamination.

accessibility data. Validity of cell type annotation to chromatin accessibility data was however checked by visually inspecting cell type annotations on the integrated ATAC UMAP, as shown in 3.10Figure B.

Initial annotation identified two clusters of ambiguous cell type, with marker gene expression from monocytes and normoblasts. These clusters also possessed a high proportion of counts removed by CellBender, as shown in Figure 2.11B, indicative of high ambient RNA contamination. These clusters were removed and data was re-integrated and clustered as described above. These clusters were again annotated with both coarse and fine cell type annotations. MACS2 was used to re-call peaks in coarse cell type annotated subsets. This improved peak calling performance for cell type specific peaks, which can otherwise be lost in rare populations.

Annotation quality control

Cluster purity was assessed by constructing heat maps of marker gene expression and chromatin accessibility. The top 5 markers for each assay were identified using presto v1.0.0 wilcoxauc command to calculate pairwise Wilcoxon test statistics between clusters. Average expression of normalised counts were calculated for each cluster and scaled feature expression was plotted using ComplexHeatmap. Validity of HSPC annotation was also checked by comparing immunophenotypic data, which was available from previous experiments, for individuals used in the 10X Multiome experiment. HSPC cells were subset from the complete multiome dataset and the percentage of each HSPC subpopulation was calculated per individual. Mean percentages of HSPC subpopulations per phenotypic group was calculated along with standard error of means for each mean percentage. Immunophenotypic data was treated similarly, mean percentage of each HSPC subpopulation was calculated per phenotypic group, along with standard error of means for mean percentages. If data for an individual was available from more than one experiment, the mean percentage between replicates was used. Stacked bar plots were created with ggplot2. Exploration of differences between multiomic and immunophenotypic proportions of HSPC subpopulation were explored by plotting mean normalised RNA expression of marker genes within each cluster combined to make each HSPC subpopulation. Percentage of cells within each phenotypic group were calculated and displayed as a stacked bar chart alongside the marker gene expression dot plot. Confirmation of HSC annotation was performed using AUCell, which calculates gene set enrichment on a per cell basis[252], and published HSC gene signatures[253], [254].

Haematopoietic stem cell subgroup analysis

The HSC cluster was subset and integration was re-run as described above on CellBender adjusted RNA and MACS2 called peak matrices, with the exception of Harmony being used to remove batch effect, rather than inter-individual variance. UMAP projections were created from each assay to visualise differences between phenotypic groups. Differential gene expression and chromatin accessibility analysis

was performed by creating aggregated pseudobulk samples for each individual and then comparing phenotypic clusters using EdgeR glmLRT. Batch effect was accounted for within the generalised linear model. Gene ontology analysis was performed using GREAT v.4.0.4 on differentially accessible peaks between phenotypic groups or between CMML and healthy donors[255]. Gene set enrichment analysis of gene expression was performed by ranking differences between phenotypic groups using presto wilcoxauc. Genes were ranked by descending auc score and fgsea was used with the 'Hallmarks' gene sets from the msigdbr v7.5.1 package. Motif enrichment analysis was performed using chromVAR to calculate deviation scores for motifs within accessible peaks for each individual cell. Transcription factor motif position weighted matrices from the JASPAR database were retrieved from the JASPAR 2020 package for this analysis. Deviation scores between phenotypic groups were then compared using presto wilcoxauc and transcription factors were ranked by auc score. Transcription factors that were important putative regulators of HSC function for each phenotype were predicted by calculating the mean auc score from differential gene expression analysis and motif enrichment analysis. Transcription factor footprinting was performed in Signac for the top putative regulators identified in the previous analysis.

Differentiation trajectory analysis

Slingshot v.2.7.0 was used to predict lineage trajectories from the combined multiome dataset using the weighted nearest neighbour UMAP projection and coarse celltype annotations to inform calculations[256]. Monocyte differentiation pseudotime and erythroid differentiation pseudotime were re-incorporated into the multiome Seurat object and visualisations for motif accessibility (chromVAR deviation score) and gene expression change over pseudotime were created using ggplot2.

FigR genome regulatory landscape analysis

The multiome dataset was split by clinical phenotype and the Functional Inference of Gene Regulations (FigR) v.1.0.1 package was used to calculate genome-wide correlations between gene expression and local chromatin accessibility to identify

domains of regulatory chromatin (DORCs)[257]. TF motif enrichment analysis was then performed on DORCs to create TF to gene links and an overall regulation score was calculated for each of these links. Global assessment of genome regulatory landscape was performed by identifying gene and TF modules controlling specific cell type differentiation programs. The elbow v.0.9 package (https://github.com/ahasverus/elbow) was used to calculate the inflection point for DORC ranked gene lists, which determined 4 to be the optimal inflection point in all phenotypic subgroups. Genes with 4 or more associated DORCs were further filtered to remove genes and TFs with <2.5 maximum regulatory score to leave the gene-TF links that were predicted to be most biologically important in the dataset. Principal component analysis was performed on regulation scores for genes and TFs separately. Ward clustering, using the hclust package, on principal components, identified clusters of genes and TFs. Gene clusters were annotated to haematopoietic cell types using over representation analysis, as described above for annotation of multiome cell types. Differences in TF activity between phenotypes were identified by subtracting absolute regulation scores for TFs from CMML phenotypes from healthy donor scores. TFs were ranked on their regulation score and plotted as a waterfall plot for each phenotype to demonstrate differences in TF activity in the strongest regulators of haematopoiesis. ComplexHeatmap was used to create heatmaps of the regulation score between TFs and genes for each phenotype. Dendrograms represent ward clustering of principal components as described above.

Mature myeloid cell subgroup analysis

Subgroup analysis was performed on mature myeloid cell populations, including cells initially annotated as being either Monocytes, Macrophages or cDC2. Similar the the HSC subgroup analysis, dimensionality reduction for RNA and chromatin accessibility assays were repeated and inter-individual variance was removed using Harmony. Harmony corrected PCA and LSI matrices were combined by wnn integration and a new UMAP projection calculated. Data were re-clustered and annotated as previously described. Cell types were re-annotated on the

basis of subgroup clusters and this was incorporated back into the complete dataset for cell-cell interaction analysis. Data integration was performed again on the mature myeloid cell subgroup, but this time with Harmony being used to remove differences for batch only to identify phenotype specific differences between individuals. Monocyte clusters of the subgroup analysis were subset and differential gene expression and chromatin accessibility analyses were performed on pseudobulks as described in HSC subgroup analysis. Tracks of chromatin accessibility were created using Signac CoveragePlot command.

Cell-cell interaction analysis

Interaction analysis was performed using multinichenetr v.2.0 [258], [259]. Cellbender adjusted RNA count data was extracted from Seurat to create a singleCellExperiment object with refined cell type annotations from mature myeloid subcluster analysis. Rare populations of HSPC were combined, making HSC/MPP and CMP/GMP groups. Dendritic cells, macrophages and B lymphocytes were excluded because there were insufficient cells in all three phenotypic groups to be used in the analysis. Individuals were assigned to three phenotypic groups: erythroid-biased CMML, myeloid-biased CMML and healthy donor. The analysis was conducted by comparing each group with the other two groups in order to identify phenotype specific interactions. Prioritisation was performed using the "lower_DE" scenario to reduce the weighting of differential expression within the calculation of the top prioritised interactions. Many of the interactions I anticipated identifying were between cell types with different abundance within each phenotypic group and therefore I did not want to bias the prioritisation algorithm against these interactions if ligand and receptor expression was not significantly different on a per cell basis. All other parameters were used at default settings. Circus plot and heatmap visualisations were created as described in the multinichenet tutorials (https://github.com/saeyslab/multinichenetr).

2.3.6 Imaging mass cytometry analysis

Imaging mass cytometry images were visualised using MCD viewer v.1.0.560.6 (Standard Biotools) to assess antibody staining performance during antibody testing and titration. Image acquisition during antibody optimisation was performed with all channels open to detect spillover. Spillover was minimal after antibody titration and therefore no compensation was required for final image analysis. Image analysis was performed using the steinbock v.0.16.3 pipeline[260]. Pre-processing was performed to remove hot pixels using the standard threshold of 50, and cell segmentation was performed in Ilastik in accordance with the ImcSegmentationPipeline[261]. Images and masks were imported into QuPath v.0.5.1[262] where I created object classifiers for all markers either with manual thresholding if there was no background staining or random forest machine learning classifiers if background was present. Mean pixel intensity and shape measurements were exported to R v4.3.1. Cell annotation was performed by creating boolean logic gates on canonical markers using the following code:

```
cell_type = case_when(
    HSPC == 1 ~ "HSPC",
    CD45 == 1 & CD15 == 1 & CD14 == 0 ~ "Mature_granulocyte",
    CD45 == 1 & CD14 == 1 & CD16 == 0 ~ "Classical_monocyte",
    CD45 == 1 & CD16 == 1 & CD163 == 0 & CD15 == 0 & CD14 == 0
        ~ "Atypical_monocyte",
    (CD45 == 1 & CD68 == 1) | (CD45 ==1 & CD163 == 1) ~ "Macrophage",
    CD45 == 1 & CSF1R == 1 & ITGB1 == 0 & CD20 == 0 ~ "CSF1R_Macrophage",
        # Negative markers added to improve purity of population
    CD45 == 1 & CD235ab == 1 ~ "Erythroid",
    CD45 == 1 & CD3 == 1 ~ "T_cell",
    CD45 == 1 & CD20 ==1 ~ "B_cell",
    CD45 == 0 & CD56 == 1 ~ "Osteocyte",
    CD45 == 0 & CD34 == 1 & aSMA == 0 ~ "Endothelial_cell",
```

```
CD45 == 0 & aSMA == 1 ~ "Smooth_muscle_cell",
CD45 == 0 & CD271 == 1 & CD14 == 0 ~ "MSC",
.default = "Other")
```

HSPC cells were identified by creating a separate random forest object classifier, to identify cells that were $CD45^{+}$ $CD34^{intermediate}$ $CD31/VWF^{-}$ $aSMA^{-}$ based on mean, max, min and standard deviation of markers. This approach allowed me to classify 42% of cells within the bone marrow. Segmentation was challenging because of the limited resolution available in imaging mass cytometry (1μm per pixel) compared to the small average size of haematopoietic cells (5-8μm in diameter) and the close proximity of these cells within hypercellular bone marrow tissue. Despite running multiple segmentation algorithms, cells were frequently found to express markers from multiple adjacent cell types. Antibody signal was also limited for anti-CD56, anti-IL1B and anti-CD123 antibodies despite using maximal titrations. However, these markers were included in the analysis because staining was deemed to be marker specific, although only the most strongly expressing cells were labelled. The purity of annotated cell types was verified by visualising scaled marker expression by cell type using ComplexHeatmap. Spatial analysis was conducted on annotated cells using the Spatial Image Analysis of Tissues (SPIAT) v.1.6.3 package to calculate distances from HSPC to all other cell types in using two orthogonal methods, as described in the SPIAT vignette[263].

2.3.7 Code availability

Code used for pre-processing, analyses and figure creation for the experiments in this thesis are available at:

https://github.com/wimm-hscb-lab-personal/Richard_Salisbury_Thesis.

3

TET2 mutation primes CMML HSPCs for myelomonocytic differentiation by hijacking the emergency myelopoiesis pathway

Contents

3.1 Introduction 72
3.1.1 HSPC function in steady state haematopoiesis 72
3.1.2 Transcription factors coordinate HSPC differentiation . 73
3.1.3 Emergency myelopoiesis and its triggers 75
3.1.4 HSPC dysfunction in CMML 76
3.1.5 Summary and plan for investigation 78
3.1.6 Aims 80
3.2 Results 80
3.2.1 HSPC populations in TET2m CMML 80
3.2.2 Patient and healthy donor characteristics of individuals included in multiomic bulk analysis 81
3.2.3 Spectrum of mutations in patient cohort 81
3.2.4 Chromatin accessibility of HSPCs separates TET2m CMML patients into three subgroups 83
3.2.5 Cluster 1 samples are enriched for SRSF2 mutations and cluster 3 samples for RAS pathway mutations 91
3.2.6 Multiomic single cell analysis reveals TET2m CMML HSCs are primed for myelomonocytic differentiation . . 91
3.2.7 Alterations to the gene regulatory landscape explain how different phenotypes develop within CMML 96
3.3 Discussion 99

3.3.1 Future work 103

3.1 Introduction

3.1.1 HSPC function in steady state haematopoiesis

Haematopoietic stem cells (HSCs) sit at the apex of haematopoiesis, from where they are responsible for repopulating all lineages of the blood by differentiation. They must also balance this with the requirement to self-renew, so that they can sustain haematopoiesis throughout the life time of the organism. As such the majority of HSC remain quiescent, with a small number giving rise to highly proliferative multipotent progenitor (MPP) populations capable of short-term self renewal and asymmetrical division[264], [265]. It is these populations that contribute to the maintenance of steady state haematopoiesis by differentiation into ever more mature progeny through multiple cell divisions. Traditionally the process of differentiation has been viewed as a tree, with lineages branching off as cells differentiate, and commitment decisions becoming more limited at each branch point. However, recent work leveraging single cell technologies has seeded doubt about this discrete hierarchical model and the reality appears more complex with conflicting interpretations within the field. Some favour a continuous model where differentiating cells can be placed on a continuous trajectory rather than at specific nodes[266] and others prefer a more deterministic model, where some HSCs are primed for differentiation to specific lineages allowing them to bypass intermediate steps. This later model has been best demonstrated in megakaryopoiesis[267].

HSCs were first discovered over 60 years ago, as cells capable of repopulating murine bone marrow after a lethal dose of irradiation[268]. Considerable research over the proceeding years led to the purification of this population, the identification of cell surface markers to aid their prospective isolation and the development of *in vitro* colony forming assays, aided by the discovery of haematopoietic cytokines[269]–[272]. These technologies, alongside genomic profiling

of immunophenotypically purified populations of HSPCs, greatly improved our understanding of the molecular mechanics underlying lineage commitment decisions with ever improving granularity[175]. This facilitated the development of a stochastic model of haematopoiesis, where lineage commitment decisions at each branch point were decided at random during steady state[273]. However, external stimulation by cytokines are capable of biasing these decisions at certain juncture when required, such as during infection or in response to cytopenia[274].

Since the turn of the century however, single cell clonogenic assays started to identify HSCs with pre-determined lineage commitment. A commitment that was heritable and therefore shared by clonally related daughter HSCs[275]. Furthermore, single cell multiomic analysis uncovered the continuous nature of haematopoiesis. Rather than discretely defined branch points, single cell analysis of gene expression and chromatin accessibility found that cells exist within a multitude of different cell states between the previously identified hierarchical states, and that the molecular dynamics of lineage commitment are constantly changing along this continuum[174]. Combining single cell genomic assays with *in vivo* lineage tracing has allowed the appreciation of heterogeneity within the HSC population, with each HSC being epigenetically primed for differentiation via a biased or balanced trajectory, with some HSCs displaying multipotency but others being restricted to differentiation to a single cell type[177], [265], [267], [276]. Most well studied of this lineage-restricted phenomenon are platelet-biased HSCs, which can exhibit megakaryocyte/platelet restricted differentiation at steady-state, but after isolation and transplantation, they retain their multipotency and are able to repopulate all blood lineages[265], [267]. Therefore, epigenetically primed HSCs retain their potential to alter their destiny when the situation dictates.

3.1.2 Transcription factors coordinate HSPC differentiation

Transcription factors orchestrate gene expression by binding to specific DNA loci. They exert their function in multiple ways, including recruitment of transcriptional machinery to cis-regulatory elements, by binding to epigenetic regulators capable of

histone modification or altering cytosine methylation, or simply by blocking other proteins from binding to the same site[277]. These functions can either be performed directly or through a co-activating intermediary protein such as CBP/p300[278]. Through these mechanisms, TFs are able to coordinate the complex changes to gene expression patterns that are required during cell differentiation. Specific TFs have been identified as being required for the development of certain cell types, such as Tal1 and Runx1 for HSC formation[279]. Ectopic expression of TFs is sufficient to reprogram cells from one lineage to another, this has been most profoundly demonstrated in the de-differentiation of mature fibroblasts to embryonic stem cells using Oct3/4, Sox2, c-Myc and Klf4[280]. Single cell multiomic profiling of haematopoiesis has allowed the characterisation of TF dynamics at unprecedented resolution, drawing together data similar to that which had previously required decades of pain-staking analyses of single TFs within a single experiment[174], [281], [282]. These technologies are leading to a rapid increase in our understanding of the molecular dynamics involved in lineage commitment and how these processes can be perturbed in disease.

Although the function and interactome of individual TFs are distinct, the process by which TFs shape lineage commitment is becoming apparent. Computational modelling of this process found that commitment occurs multiple generations before lineage specific protein marker expression becomes detectable[283]. Furthermore, epigenetic priming for specific TF function precedes its detectable RNA expression, although gene expression is able to detect lineage priming when cell type specific genes are expressed by progenitors[174], [284]. Multiple mechanisms are involved in epigenetic priming, but each of these converge on the alteration of chromatin accessibility, which can be detected by assay for transposase-accessible chromatin with sequencing (ATAC-seq) or DNase-seq. Transcription factor binding site enrichment within areas of open chromatin can be used to accurately predict fate restriction in HSCs, particularly when enrichment is found within enhancer elements as opposed to promoters[276]. Therefore, epigenetic profiling offers the optimal technique to detect lineage priming. However, because chromatin accessibility can

be altered in multiple ways, further downstream investigation is required to identify mechanisms responsible for these differences.

3.1.3 Emergency myelopoiesis and its triggers

Haematopoiesis can rapidly adapt to external stimuli to help protect an organism from attack. Microbial infection or tissue damage during trauma can trigger emergency myelopoiesis (EM), which ramps up production of myelomonocytic innate immune cells, bolstering the body's first line of defence against microbes. At the same time EM down-regulates erythroid and lymphoid cell production, which can result in anaemia, such as that seen in critical illness. HSPCs are sensitive to a vast array of stimuli, such as changes in hormone expression, cytokine gradients, damage-associated molecular patterns (DAMPs) and pathogen-associated molecular patterns (PAMPs)[285]–[287]. During steady state, the majority of HSCs are quiescent, but external stimulation leads to their activation, prompting them to switch metabolic state, transitioning from anaerobic glycolysis to oxidative phosphorylation and enter the cell cycle[288]. Proliferating HSCs form short term multipotent progenitors (MPP), which retain some self renewal capacity. MPP in mice have been subclassified into 4 distinct functional groups, where MPP3 are myeloid-biased[284]. MPP3 in EM function as a reservoir for rapid granulocyte-macrophage progenitor (GMP) production and a subset of MPP3 develop a secretory phenotype, releasing pro-inflammatory cytokines including TNFα and IL-6, which further amplify myeloid cell production by HSPCs[289]. GMPs rapidly proliferate and form clusters, creating a production line for myelomonocytic cells[290]. Cytokine release can also influence the function of other HSPCs; IL-6 reprograms multipotent lymphoid-biased progenitors to become exclusively myeloid-biased[291], IL-33 inhibits erythroid progenitor differentiation and IL-22 increases erythroid progenitor apoptosis[292], [293].

EM has been extensively modelled using various stimuli in the attempt to elucidate the mechanisms by which HSPCs transition from steady state to myeloid-biased haematopoieis. Lipopolysaccharide (LPS) treatment of common myeloid progenitors (CMP) leads to reduced expression of TFs *Gata1, Bach1* and *Bach2*,

with increased *Cebpb, Csf1r* and *Spic*[294]. Bach 1/2 share genomic binding site loci with C/EBP-β proximal to genes involved in myeloid differentiation and function[294]. BACH1/2 and C/EBP-β can therefore act as a logic switch by competing against each other to define myeloid or erythroid lineage fate[295].

C/EBP-α and PU.1 are the most important regulators of steady state myelopoiesis. EM however relies on the function of C/EBP-β and is entirely independent of C/EBP-α activity[296]. Interestingly C/EBP-α and C/EBP-β are highly similar. They share conserved C-terminal regions and sequence similarities in TF binding site preference. Inserting *Cebpb* into the *Cebpa* locus in *Cebpa* KO mice rescues granulopoiesis[297]. However, they also have some important differences. C/EBP-β exerts a stronger proliferation stimulus to C/EBP-α by de-repression of c-Myc[296]. C/EBP-β also performs an essential role in innate immune memory[298]. Innate immune memory occurs within HSCs that have been exposed to an inflammatory stimulus such as LPS. This creates a heritable epigenetic environment priming the HSCs and their daughter cells to myelomonocytic differentiation in response to a secondary infection by maintaining open chromatin at C/EBP-β target loci by an as yet unknown mechanism.

3.1.4 HSPC dysfunction in CMML

HSPC dysfunction has been demonstrated to be central to CMML pathogenesis. Altered *in vitro* colony forming potential was one of the first observations made by those investigating CMML, leading to the suggestion that this was a leukaemia developing within stem cells almost 50 years ago[299]. Culture of peripheral blood mononuclear cells in semi-solid methylcellulose gives rise to a skewed colony output with defective erythroid burst (BFU-E) and erythroid colony formation (CFU-E) and enhanced production of granulocyte-macrophage colonies (CFU-GM)[300]. Immunophenotyping and *in vitro* culture assays on prospectively purified HSPC sub-populations led to the discovery that MPP and CMP populations are expanded in CMML, whereas GMP populations are quantitatively and qualitatively normal. MPP and CMP exhibit enhanced CFU-GM formation in semisolid

culture and skewed granulomonocytic/erythroid output in single cell liquid culture assays[16]. These findings demonstrate an inherent myelomonocytic differentiation bias within these multipotent progenitors, which correlates with the patient's clinical phenotype[16], [301]. Multiple iPSC models have also been created that recapitulate the CFU findings from primary patient samples[302]–[304]. These models benefit from being able to isolate multiple clones from the same patient and investigate them separately to identify the effect of specific genotypes on colony forming potential. KRAS mutations for instance are associated with increased monocyte production in liquid culture assays, larger colony size and increased CFU-M (macrophage)[302]. These data clearly demonstrate that multipotent CMML HSPCs are primed for myelomonocytic differentiation and provide evidence that *in vitro* culture assays are a relevant model for investigating CMML HSPC dysfunction.

Single cell analysis of primary CMML patient HSPCs has provided some insight into the potential mechanisms priming HSPC for altered function. Multipotent $CD34^+$ $CD38^-$ CMML HSPCs are transcriptomically distinct from healthy donor HSPCs. Upregulated genes in CMML are enriched within cell-cycle and myeloid immune cell pathways such as cytokine-cytokine receptor interaction and JAK-STAT signalling[305]. *FOS* and *JUN*, activating protein-1 (AP-1) family members, were consistently down regulated in CMML samples. Gene set enrichment analysis, using published signatures of oligopotent HSPCs, found enrichment of neutrophil primed and monocyte/dendritic primed progenitors but negative enrichment for HSC markers, indicating that even at this early stage, HSC gene expression is already showing signs of lineage bias[305]. These findings have been backed up by a recent scRNA-seq analysis of $CD34^+$ HSPCs from 39 CMML patients, which used the Palantir package to construct probability maps of HSPC differentiation. This found, somewhat surprisingly, that CMML patients can be divided equally into three separate phenotypic groups - monocyte-biased, MEP-biased and normal-like; based on their altered cell densities along differentiation pseudotime[17]. Although those patients in the monocyte-biased group had the strongest predicted bias to this differentiation trajectory, most patients in the other phenotypic groups

also had greater probability of differentiation towards monocytes than controls. Monocyte-biased CMML patients were associated with increased blast count and RAS pathway mutation. The term 'blast' refers to an aberrant clonal population of HSPC-like or progenitor-like cells, which possess self-renewal potential and can often be leukaemogenic if transplanted. Blasts detected in monocyte-biased CMML were frequently CD120b^{+} GMP-like blasts, which were highly inflammatory and transcriptionally similar to expanded cluster forming GMP cells identified in EM[289], [290]. This raises the possibility that similar epigenetic or signal transduction pathways may be utilised to promote myelomonocytic expansion in CMML as those used in EM.

Advanced stage CMML also possesses distinct HSPC features, with a negative correlation being observed between blast percentage and the size of HSC populations[17], [306]. Blasts within chronic phase CMML can be similar to multiple different HSPC populations but the molecular determinants of this remain to be discovered. As discussed above, monocyte-biased CMML phenotype is associated with a GMP-like blast population, which becomes further expanded by administration of LPS in murine xenograft models of CMML[17]. Therefore, determinants of blast phenotype may be cell-intrinsic or cell-extrinsic in nature. Similar to blasts detected in chronic phase CMML, transformation to secondary AML, identified by the presence of an expanded blast population contributing >20% of total bone marrow nucleated cells, can arise within oligopotent progenitors as well as HSC/MPP[94], [306]. Oligopotent blasts can de-differentiate into an HSC-like state, upregulating HSC specific gene programmes, and presumably enhancing self-renewal potential of the clone[306]. Therefore, blasts at the point of transformation likely possess different function to blasts or malignant HSPCs in chronic phase CMML. I have therefore chosen to study patients in chronic phase CMML rather than after transformation.

3.1.5 Summary and plan for investigation

TET2 loss of function mutations (TET2m) are usually the founding mutation of a CMML clone[16]. Loss of TET2 function leads to increased methylation, particularly

at enhancer elements and gene bodies where methylation is no longer being actively removed by TET2[156]. TET2 loss also interferes with the function of associated histone modifiers that require TET2 to form a chromatin regulatory complex at specific genomic loci[161], [162]. TET2 KO murine studies found that the catalytic function is most important for regulating myelopoiesis but the non-catalytic role in facilitating histone modification is important for enhancing HSPC self-renewal[171]. Both of these mechanisms are therefore important in CMML pathogenesis, but the specific effect on gene expression will be difficult to detect without a comprehensive screen of changes to methylation and a wide array of histone modifications, which is particularly challenging when using primary patient samples where cell numbers are limited. In addition, methylation change is heritable and therefore changes in methylation may occur in an immature cell but not exert its CMML promoting effect until a later stage of development confounding the interpretation of differentially methylated regions in progenitor populations. I therefore chose instead to use ATAC-seq to identify loci with altered chromatin accessibility to provide insight into the likely role of TET2m on altering CMML HSPC function. Regardless of the specific mechanism altering chromatin conformation, by linking the epigenetic change to gene expression, I am able to infer the paths through which TET2 loss of function can contribute to CMML pathogenesis, leaving the specific mechanism underlying this change to be elucidated by future work. I couple this with an analysis of gene expression using RNA-seq to provide insight into how HSPC function changes in TET2m CMML. Bulk level analysis however can only provide information on the greatest differences because subtler changes that are lost due to the noise created by cellular heterogeneity within the sample tested. Therefore, to gain greater granularity on these differences I also perform single cell multiomic analysis on a subgroup of patients and healthy donors where it is possible to infer greater links between changes in chromatin conformation and gene expression. By doing so I identify cell type specific and global changes to gene regulatory networks that provide a new understanding of the pathogenesis of TET2m CMML.

3.1.6 Aims

The specific aims for this section of my thesis are:

1. To characterise the spectrum of HSPC dysfunction in TET2 mutant CMML
2. To explore epigenetic drivers of phenotypic divergence in TET2 mutant CMML HSPCs
3. To identify HSPC phenotype specific alteration of gene regulatory networks responsible for TET2 mutant CMML pathogenesis

3.2 Results

3.2.1 HSPC populations in TET2m CMML

TET2 loss of function alters HSC function in murine knockout models and human *in vitro* systems, skewing differentiation towards myelomonocytic lineages and enhancing HSC self-renewal leading to the accumulation of HSC and CMP subpopulations [164]–[167], [169], [179]. To explore the effect of TET2 mutation in the context of CMML, I performed immunophenotyping on lineage negative $CD34^+$ HSPCs from 53 cryopreserved bone marrow MNC or G-CSF mobilised peripheral blood samples. Multiple samples from the same individual were combined by using the mean proportion of each subpopulation from the samples assayed. Data were available from 23 TET2m CMML patients and 15 healthy donors. Surprisingly, TET2m CMML patients had reduced CMP and increased MEP as a proportion of total HSPCs than healthy donors (Figure 3.1A). There was also a trend towards a decrease in HSCs (p=0.052). Variance of oligopotent progenitors (CMP, GMP and MEP) was greater in TET2m CMML patients than in healthy donors. To explore this further I correlated HSPC populations against one another and found that in patients, but not in healthy donors, CMP and GMP populations were negatively correlated (r=-0.49, p=0.004; Figure 3.1B). MEP were also negatively correlated with CMP and GMP in both patients and healthy donors. In keeping with this correlation analysis, no patients had both elevated CMP and GMP proportions,

but some patients had reduced CMP and GMP proportions. This group of patients possessed an expansion of MEP (MEP proportion >0.2 of total HSPC), which allowed patients to be divided roughly into three groups, those that had either increased CMP, GMP or MEP proportions as shown in Figure 3.1C. These findings were not in keeping with murine TET2 KO models, which may be due to differences between murine and human haematopoiesis or the contribution of co-mutations altering the effect of TET2 mutation in CMML.

3.2.2 Patient and healthy donor characteristics of individuals included in multiomic bulk analysis

To understand the potential drivers of the observed heterogeneity in TET2m CMML HSPCs, I conducted multiomic profiling of Lin^- $CD34^+$ HSPC populations from 18 TET2m CMML patients and 10 healthy donors. Within this cohort, CMML patients were older than healthy donors (mean 75.1 vs 55.8yrs, p=0.01) and a higher proportion were men (83% vs 44%, p=0.07). A summary of CMML patient demographics are displayed in Table 3.1.

3.2.3 Spectrum of mutations in patient cohort

CMML patients had a median of 4 detected mutations (95% CI 3.42-4.47) as shown in Figure 3.2A. 30 TET2 mutations were detected in the 18 CMML patients, with 1 mutation being classified as a variant of uncertain significance (VUS). This mutation, TET2:NM_001127208:exon4:c.3499A>G:p.R1167G, was highly clonal and associated with loss of heterozygosity resulting in a variant allele frequency of 90%. Although this substitution has not been previously implicated in haematological malignancy, it has been reported in ovarian cancer and therefore was deemed likely pathogenic so this patient was included in subsequent analyses[307]. TET2 mutations were detected throughout the length of the protein, with the majority (18/30, 60%) resulting in a truncation of the coded protein and the remaining mutations resulting in a substitution (12/30, 40%). Substitutions were predominantly found at either end of the TET2 catalytic domain (Figure 3.2B),

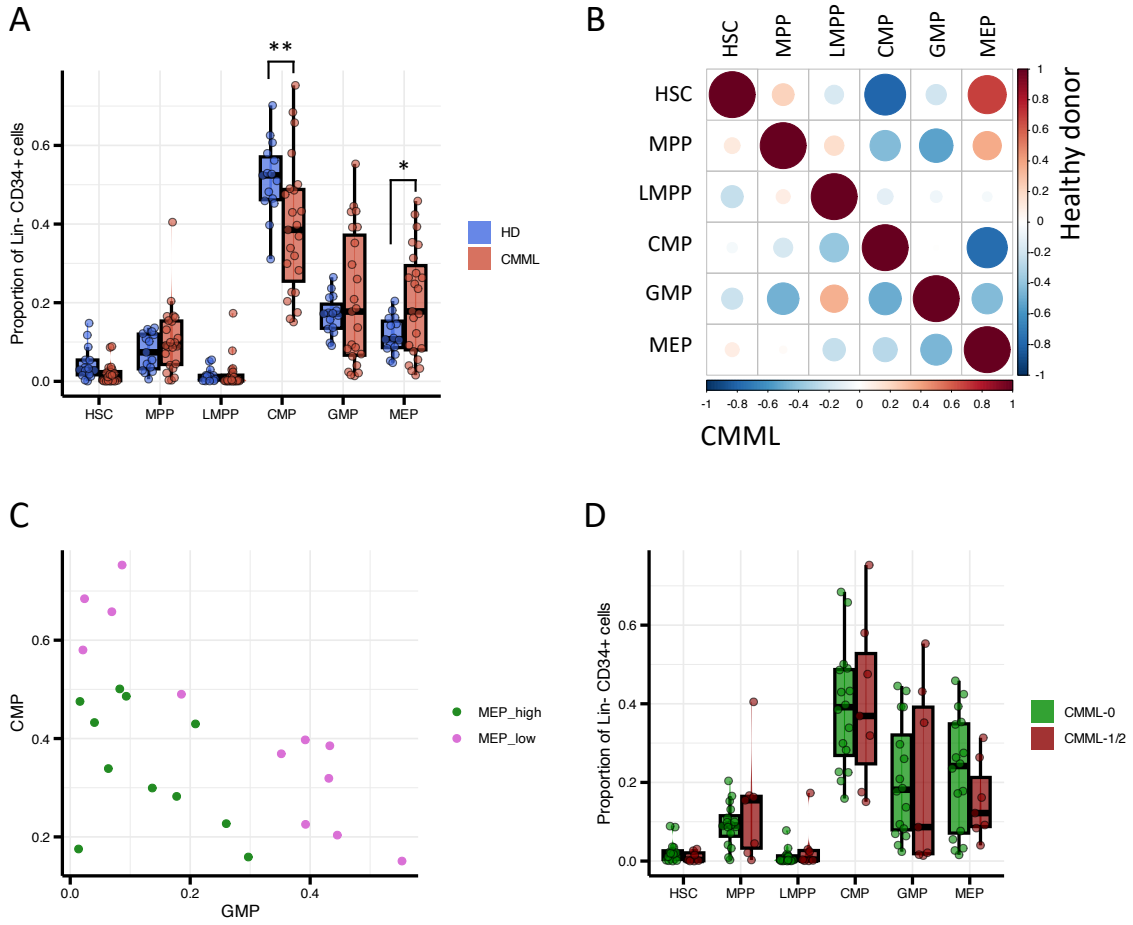


Figure 3.1: HSPC proportions are altered in TET2m CMML - (A) Immunophenotyping was conducted on Lin$^-$ CD34$^+$ HSPCs from 23 TET2m CMML patients and 15 healthy donors. Data displayed split by HSPC subpopulation. Proportion of total HSPC compared between healthy donor and CMML patients by Wilcoxon test. *p<0.05, **p<0.01. (B) Pearson correlation coefficient between HSPC subpopulations in healthy donors (top right) and TET2m CMML patients (bottom left). Colour of each circle denotes corresponding correlation coefficient and size inverse p value. (C) Comparison of CMP and GMP proportions in TET2m CMML patients. Each point represents an individual and those with MEP proportion of total HSPCs >0.2 are highlighted in green. (D) Comparison of HSPC subpopulations between CMML-0 (n=16) and CMML-1 or -2 patients (n=7)

Table 3.1: TET2m CMML patient demographics

	TET2m CMML (n=18)
Age	75.1yrs (SD 4.6)
Sex (male)	15 (83.3%)
Peripheral blood counts	
Hb	120.1g/L (SD 23.9)
WCC	21.7x10^9/L (SD 20.8)
Neut	12.2^9/L (SD 13.6)
Mono	6.0^9/L (SD 4.8)
Plt	94.2^9/L (SD 52.4)
Phenotype	
MD-CMML	6 (33.3%)
MP-CMML	12 (66.7%)
Stage (WHO 2016 classification)	
CMML-0	12 (66.7%)
CMML-1	5 (27.8%)
CMML-2	1 (5.5%)
CPSS-Mol	
Low	2 (11.1%)
Int-1	3 (16.7%)
Int-2	12 (66.7%)
High	1 (5.5%)

similar to those reported in the pan-cancer cBioPortal database, shown in Figure 1.2A. In total 17/18 (94%) patients had predicted biallelic TET2 loss of function, with ≥2 TET2 mutations detected in 13 patients. 4 further patients had a single TET2 mutation with VAF >55%, suggestive of copy neutral loss of heterozygosity, with normal karyotype.

3.2.4 Chromatin accessibility of HSPCs separates TET2m CMML patients into three subgroups

ATAC-seq libraries were successfully created from Lin$^-$ CD34$^+$ HSPCs from 15 TET2m CMML patients and 9 healthy donors, summarised in Figure 3.3B. Principal component analysis of log2 CPM normalised peak counts separated TET2m CMML samples from healthy donor samples along the PC2 axis (Figure 3.3C). TET2m CMML samples were also divided into 3 separate groups along the PC1 axis, suggesting that ATAC-seq was successfully able to identify variance between patients.

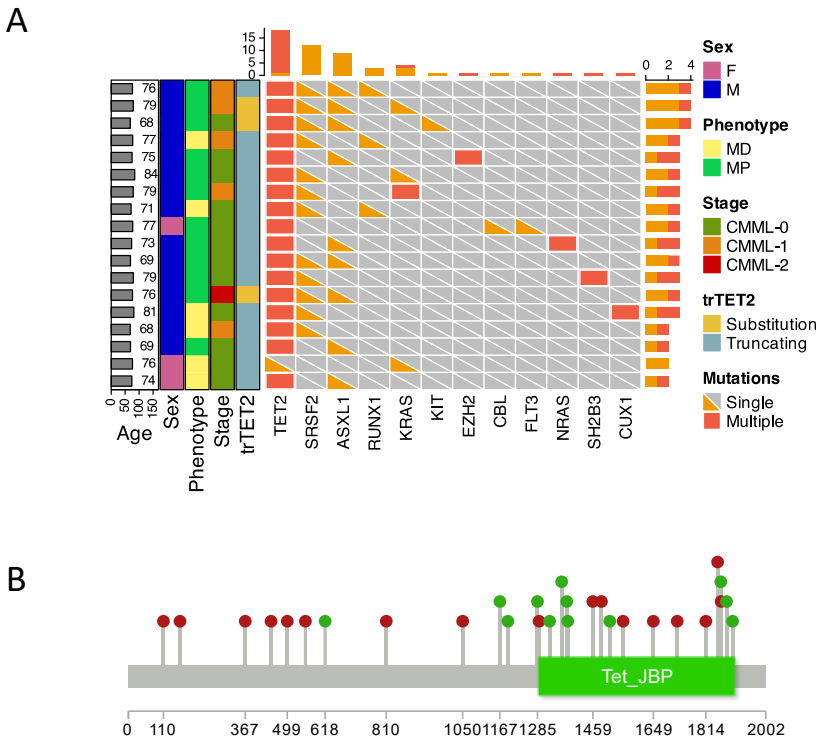


Figure 3.2: Somatic mutations of CMML patients used in bulk assays - (A) Summary of somatic mutations identified in CMML patients and associated demographic details. Patients are represented by rows and mutations/demographics by columns. Stage was classified according to the 2016 WHO classification of myeloid neoplasms[308]. Truncating TET2 mutations, trTET2, are either nonsense or frameshift mutations that introduce a premature stop codon, substitutions are either missense mutations or in-frame insertions/deletions. Multiple mutations are either $\geq$2 mutations detected or variant allele frequency of a single mutation >55%. (B) Lollipop plot of locations of individual TET2 mutations. Scale refers to amino acids in protein. Substitutions and truncating mutations are displayed in green and red respectively. TET_JBP is the catalytic domain of TET2.

Unsupervised Ward clustering of the first 23 principal components identified 4 clusters: 2x TET2m CMML clusters (cluster 1 and 2, green and pink respectively), 1x healthy donor cluster (cluster 4, yellow) and a mixed cluster with 3x TET2m CMML and 2x healthy donor samples (cluster 3, blue; Figure 3.4A). Differential accessibility analysis found that cluster 1 had 9795 peaks with increased accessibility and 2307 peaks with reduced accessibility compared to all other clusters. Cluster 2 however had 2675 peaks with increased accessibility and 7629 with decreased accessibility. This difference in distribution suggests that TET2 mutation may be having a different effect on chromatin accessibility in samples within these two clusters.

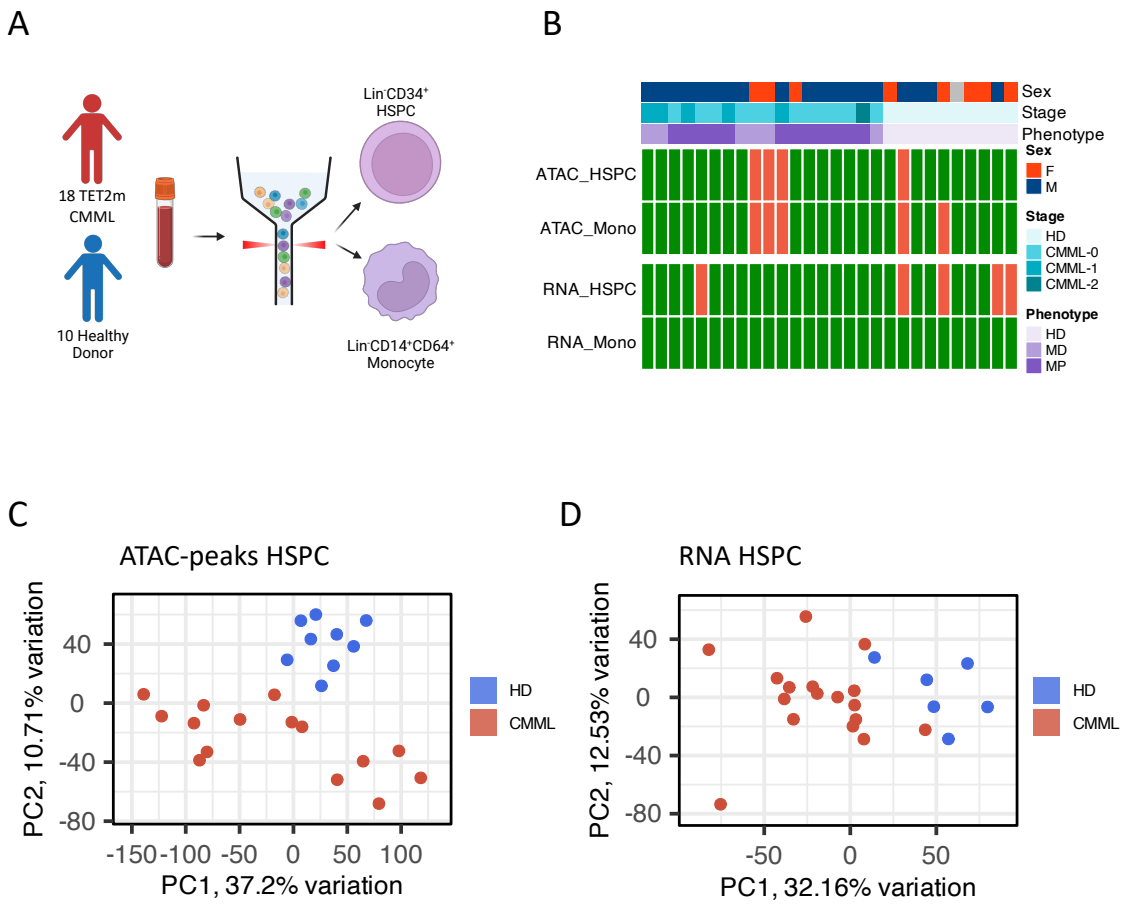


Figure 3.3: Multiomic analysis of HSPCs identifies three subgroups within TET2m CMML patients - (A) Overview of multiomic analysis of sorted bulk populations of HSPCs and monocytes. (B) Heatmap of assays included in analysis. Successfully sequenced libraries are in green and failed libraries are in red. (C) PCA of TMM normalised log2 CPM of peaks from ATAC-seq samples. TET2m CMML are in red and healthy donor samples in blue. (D) PCA of batch corrected TMM normalised log2 CPM RNA-seq samples. Colours correspond with the same groups described in (C).

TET2 loss of function is associated with global hypermethylation of DNA[167]and would therefore usually be expected to be associated with reduced chromatin accessibility[156]. To explore the probable effects of these chromatin accessibility changes, I performed gene ontology analysis using GREAT v4.0.4 [255] on peaks that were differentially accessible in each cluster (Figure 3.4E and F). Open peaks in cluster 1 were predicted to regulate genes involved in erythrocyte differentiation and mitochondrial function, whereas open peaks in cluster 2 were involved in myeloid cell differentiation and function, as well as carbohydrate and phospholipid metabolism.

These findings were suggestive that clusters 1 and 2 are functionally distinct and primed for differentiation into erythroid and myeloid lineages respectively.

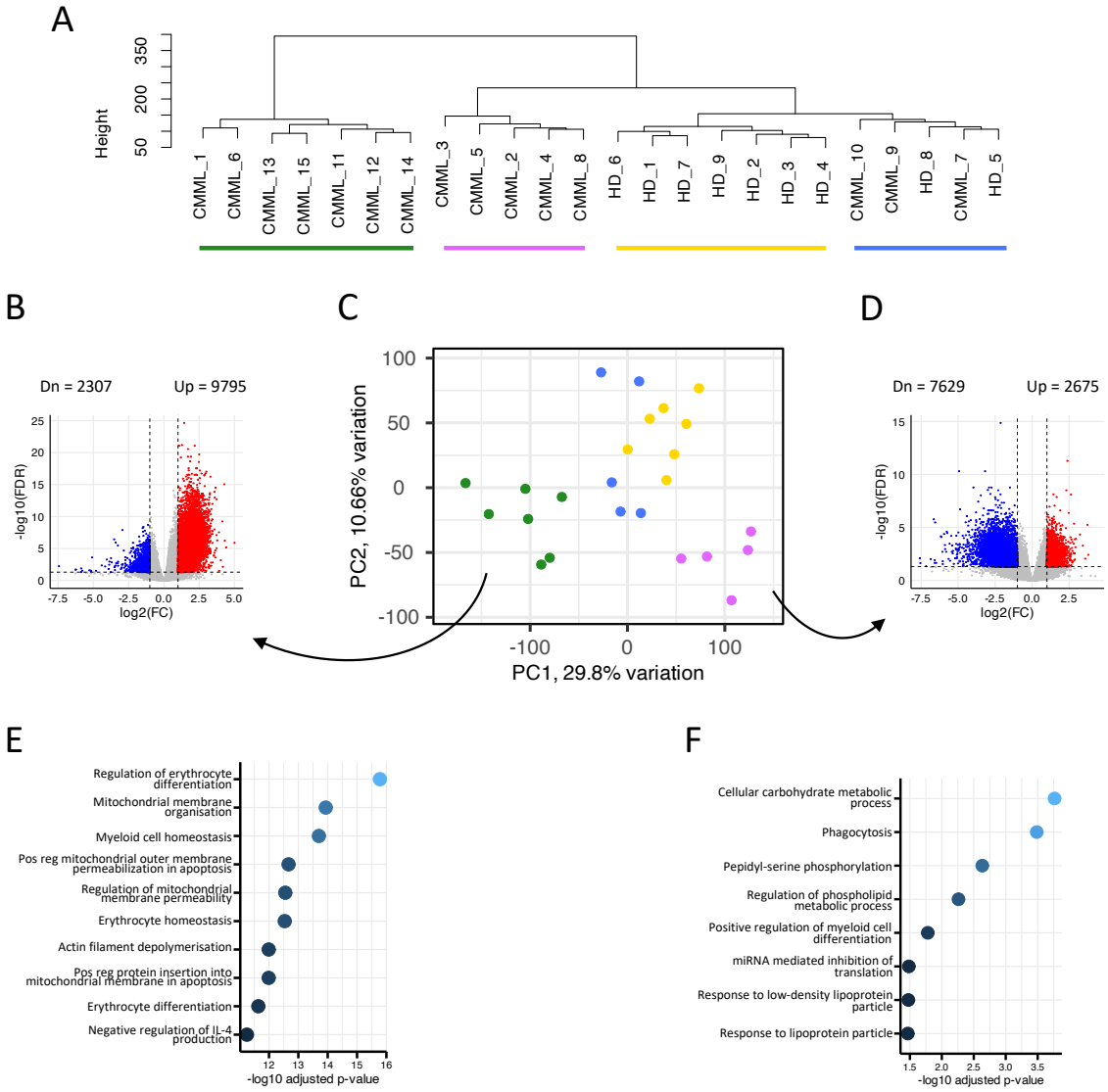


Figure 3.4: Chromatin accessibility analysis identifies distinct phenotypic groups within TET2m CMML HSPCs - (A) Dendrogram of Ward clustering of peak counts. Clusters 1 to 4 are highlighted in green, pink, yellow and blue respectively. (B) Volcano plot of differentially accessible peaks between cluster 1 (green) and all other clusters. (C) Clusters from (A) overlayed on PCA plot. (D) Volcano plot of differentially accessible peaks between cluster 2 (pink) and all other clusters. (E) GREAT gene ontology enrichment of significant open peaks from cluster 1 (E) and cluster 2 (F)

To investigate this further I performed motif enrichment analysis using chromVAR[223]. chromVAR calculates a deviation score, which represents the scaled difference in enrichment of TF motifs within open chromatin in each sample. Pearson

correlation analysis of deviation scores found strong correlation between samples in cluster 1 and also between samples in cluster 2. However some healthy donor samples switched from being in cluster 4 to cluster 3, demonstrating that heterogeneity exists within healthy donor samples as well as between TET2m CMML patients, creating a group of highly similar healthy donors in cluster 4 (Figure 3.5A). Samples in cluster 3 however were more heterogeneous, as demonstrated by weaker correlation between samples, than those in the 3 other clusters. TF motifs differentially enriched in each cluster were determined by Wilcoxon test between the cluster of interest and all other clusters. This found GATA1 and TAL1 motifs to be highly enriched in cluster 1, TFs that are essential for coordinating erythropoiesis[309]. Cluster 2 was highly enriched for C/EBP family motifs, suggestive of a myeloid phenotype. Cluster 4 was enriched for motifs associated with AP-1 complex members, the FOS-like and JUN-like families. AP-1 can be composed of multiple FOS, JUN and ATF family members that form a complex capable of binding AP-1 motifs on DNA. The combination of TFs forming the complex can have markedly different functional effects and therefore predicting the likely impact of these motifs being enriched within cluster 4 is challenging[310]. However some studies have found c-FOS and JUNB negatively regulate haematopoietic stem cell proliferation [311], [312], suggesting that this cluster may have a more quiescent, stem cell-like phenotype.

RNA-seq analysis supports phenotypic division of TET2m CMML HSPCs into 3 groups and uncovers shared alterations in HSPC function

To investigate the downstream effects on transcription of altered chromatin accessibility in TET2m CMML I performed RNA-seq analysis using SmartSeq3 chemistry on sorted Lin^- $CD34^+$ HSPCs from 17 TETm CMML and 6 healthy donors [213]. 14 TET2m CMML and 6 healthy donor samples had paired ATAC-seq data available as outlined above (Figure 3.3B). Batch effect was corrected using limma[231] and PCA analysis on log2 normalised CPM found HSPC samples separated between TET2m CMML samples and healthy donors on PC1 axis, but the clusters identified using ATAC-seq were not concordant with clustering of RNA-seq data (Figure 3.6A and B). However, gene set enrichment analysis (GSEA), using curated gene sets of

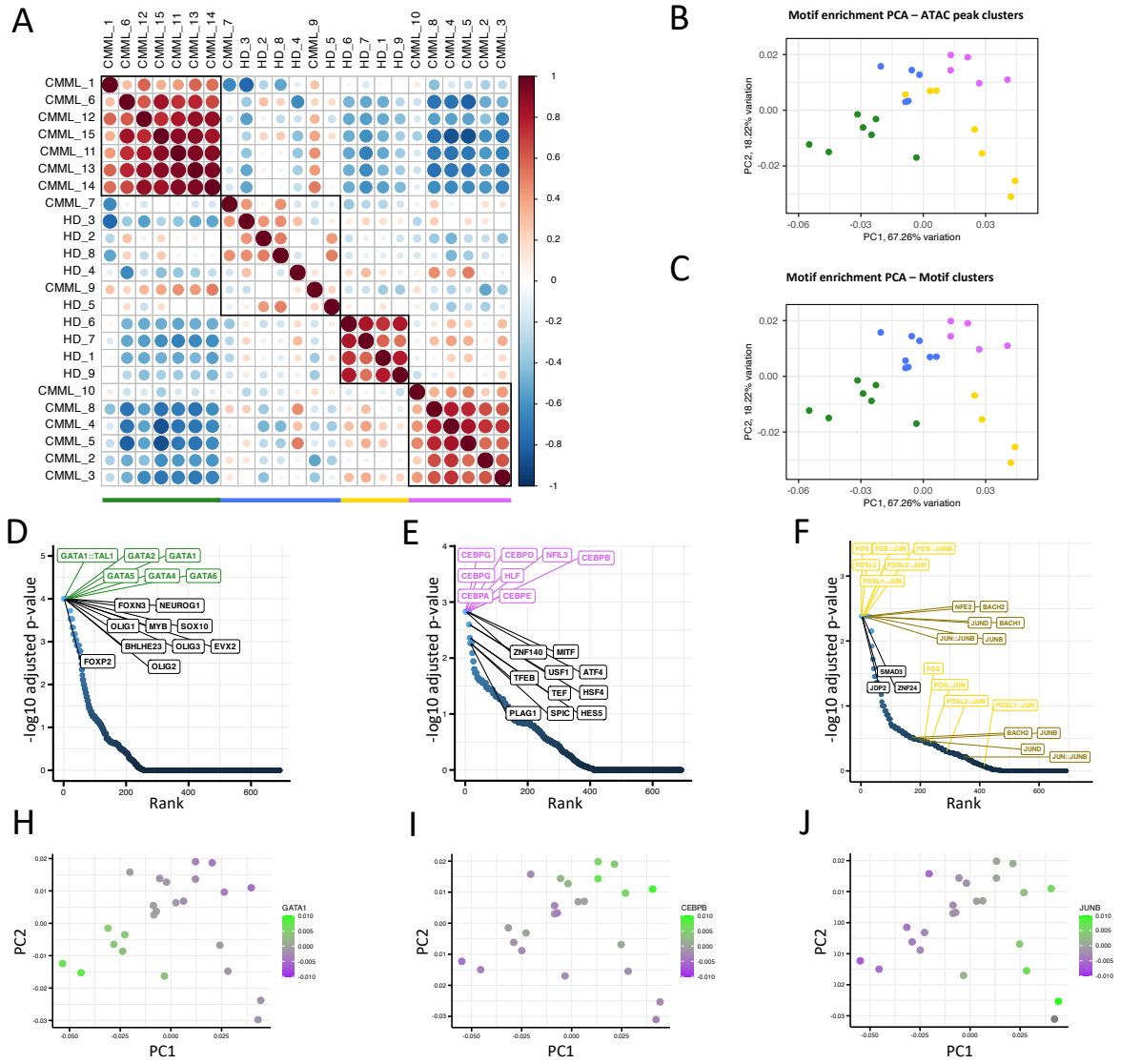


Figure 3.5: Motif enrichment analysis identifies phenotypic differences between TET2m CMML subgroups - (A) Pearson correlation analysis of motif enrichment deviation scores between samples. Clusters 1 to 4 are highlighted in green, pink, yellow and blue respectively. PCA analysis of motif enrichment deviation scores coloured by clusters identified from PCA analysis of peaks (B) and Ward clustering on PCA of motif enrichment deviation scores (C) demonstrating changes to individuals in clusters 3 and 4. Differences in motif enrichment scores between clusters were detected by Wilcoxon test and TF motifs were ranked by area under the curve statistic. TF motif enrichment specific for each cluster is displayed for cluster 1 (D), cluster 2 (E) and cluster 4 (F). GATA family motifs are displayed in green, C/EBP family in pink, FOS-related family in yellow and JUN-related family in gold. PCA plots for top ranked TF motifs enriched in cluster 1 (H), cluster 2 (I) and cluster 4 (J) coloured by deviation score.

bulk sorted and single cell haematopoietic reference data sets, found that erythroid progenitor genesets were enriched within cluster 1 samples and myeloid progenitor

gene sets within cluster 2 (Figure 3.6C and D respectively)[313]–[315]. This is highly suggestive of altered HSPC priming between these two clusters, and/or differences in the proportion of progenitor subpopulations being found within these samples. These gene expression differences are likely to be driven by TFs with binding motifs enriched within open chromatin, such as GATA1, promoting erythroid differentiation, and CEBPA, promoting myelomonocytic differentiation, in cluster 1 and 2 respectively. In support of this, gene expression of *GATA1* (p=0.008) and *CEBPA* (p=0.042) are increased in cluster 1 and 2 respectively, compared to each other (Figure 3.6E). Phenotypic differences may also be driven by downregulation of TF activity, such as BACH2, which represses myeloid gene expression[316]. *BACH2* expression was reduced in TET2m CMML clusters compared to healthy donor cluster 4 (p=0.032).

Although understanding the drivers of these phenotypic differences are undoubtedly important for disentangling the pathogenesis of TET2m CMML, PCA analysis of RNA-seq data found that the greatest source of variance were differences between TET2m CMML and healthy donor samples rather than between TET2m CMML phenotypic clusters. I therefore performed differential expression analysis using DESeq2 and found that 45 genes had increased expression and 57 genes has decreased expression between TET2m CMML and healthy donor samples (alpha = 0.05; Figure 3.6F). GSEA analysis using MSigDB hallmark genesets found enrichment for MYC targets and E2F targets, suggestive of TET2m CMML HSPCs being more proliferative than healthy donor HSPCs, and oxidative phosphorylation and fatty acid metabolism gene sets. Alterations in metabolism are critical to haematopoietic stem cell (HSC) function. HSCs reside in a hypoxic environment where they rely on glycolytic metabolism whilst quiescent, but switch to fatty acid metabolism when proliferating without differentiation and oxidative phosphorylation during differentiation[288]. Unexpectedly gene sets involved in interferonγ response and TNFα signalling via NFκB were enriched in healthy donor HSPCs. This is at odds with some published work on HSPCs in CMML, where HSPCs were found to be enriched for pathways involved in inflammation [305], [317], but not others on TET2m CHIP, where downregulation of inflammatory response pathways was

implicated in promoting clonal fitness in an inflammatory environment that could otherwise promote stem cell exhaustion[48].

Re-analysis of immunophenotyping finds that chromatin accessibility changes are associated with altered HSPC composition

Clustering of samples using chromatin accessibility identified 4 clusters with different phenotypic properties. Gene ontology analysis and motif enrichment analysis identified that HSPCs in different clusters were primed for differentiation into specific lineages. These same phenotypic differences, most prominently for erythroid differentiation in cluster 1 and myeloid differentiation in cluster 2, were confirmed using RNA-seq gene set enrichment analysis. Each sample assayed is made up of RNA and DNA taken from a few thousand HSPCs, therefore one potential explanation for the observed findings is a change in the proportion of lineage committed oligopotent HSPCs present in each sample. To confirm differences in the composition of HSPC subpopulations within samples in each cluster, I re-analysed the flow cytometry data collected when sorting these samples. This confirmed that there were some important differences in the HSPC subpopulation composition between different clusters. The proportion of HSCs within Lin^- $CD34^+$ HSPCs was reduced in cluster 2 samples compared to healthy donor cluster 4 ($p=0.015$), with a trend for reduction in cluster 1 samples ($p=0.109$), shown in Figure 3.7B. MEP proportion was increased in cluster 1 compared to both the healthy donor cluster 4 ($p=0.033$) and cluster 2 ($p=0.03$), shown in Figure 3.7C. Variance of GMP proportions was particularly large in cluster 2. As previously shown TET2m CMML patients either have increased CMP or GMP proportions. Combining these two populations reduced the variance between samples in cluster 2, and there was a trend towards decreased CMP or GMP proportion in cluster 1 ($p=0.091$). Therefore, the molecularly defined erythroid bias of cluster 1 samples can be explained by the increase in MEP populations found in these individuals. However, there was no similar increase in CMP or GMP found in cluster 2 individuals. The myeloid priming of cluster 2 samples is therefore not down to the expansion of myeloid restricted oligopotent progenitors but is likely to be a more subtle shift in differentiation bias of

all HSPCs towards myeloid differentiation at the expense of erythroid differentiation. However, such a shift can only be confirmed by single cell profiling.

3.2.5 Cluster 1 samples are enriched for SRSF2 mutations and cluster 3 samples for RAS pathway mutations

To explore the clinical correlates with the phenotypic clusters, I performed a univariate Kruskal-Wallis analysis and found that SRSF2 mutation was enriched in cluster 1 samples (OR 4.33, 95% CI 1.44-13.03, p=0.009) and RAS pathway mutations (KRAS, NRAS or CBL) were enriched in cluster 3 samples (OR 23, 95% CI 3.25-162.98, p=0.002) as shown in Figure 3.8A. Univariate analysis of age, peripheral blood counts (Haemoglobin concentration, white cell count, monocyte count and platelet count), ASXL1 mutation, myeloproliferative/dysplastic phenotype and CPSS-mol category were not significantly different between phenotypic clusters within TET2m CMML. Kaplan-Meier analysis, shown in 3.8B, found no significant difference between phenotypic clusters in survival, although the analysis was underpowered due to small sample numbers in each group. Multivariable Cox proportional hazard regression analysis also found no specific factor was significantly associated with survival, although there was a trend towards patients with a myeloproliferative phenotype having a worse prognosis and patients with a truncating TET2 mutation (trTET2) having an improved prognosis (Figure 3.8C). The detection of genotype enrichment within specific phenotypic clusters suggests that CMML pathogenesis may occur via different pathways in patients within different clusters, however caution must be taken when interpreting these findings given the small numbers of samples in each group.

3.2.6 Multiomic single cell analysis reveals TET2m CMML HSCs are primed for myelomonocytic differentiation

Multiomic single cell profiling of TET2m CMML

TET2m CMML patients can therefore be subdivided into three phenotypic categories based upon multiomic profiling of their HSPCs, with molecular correlates that hint at distinct biology between these groups. HSPCs however are a group of highly

heterogeneous cell types, continuously differentiating from HSC to oligopotent progenitor cells. Bulk analysis is excellent at identifying changes to the proportion of sub-populations within the HSPC umbrella, as demonstrated by the correlation with immunophenotypic changes. However the differences between individual cell types, especially rare populations like the HSCs, are lost. Therefore in order to explore these differences in greater resolution I selected 9 representative individuals for multiomic single cell profiling. These were chosen to represent erythroid-biased CMML (cluster 1, n=3), myeloid-biased CMML (cluster 2, n=3) and healthy donors (n=3, summarised in Figure 3.9A). Individuals were selected using either immunophenotypic analysis or molecular profiling as described above. Demographics for these individuals are summarised in Figure 3.9B. Bone marrow mononuclear cells were sorted to remove dead cells and $CD34^+$ HSPCs were enriched, aiming to represent 50% of the captured cells. Cells were assayed to create paired snRNA-seq and scATAC-seq libraries from each individual cell using the 10X Genomics Multiome platform. Following genetic demultiplexing, correction for ambient RNA contamination and doublet removal, 64423 cells passed both RNA-seq and ATAC-seq quality control filtering. These cells included an even representation between individuals (mean 7158 cells per individual, inter-quartile range 6059-8257) and phenotypes (mean 21474 cells, inter-quartile range 18079-24869). Data were integrated to remove inter-individual variance from both transcriptional and chromatin accessibility data using Harmony[245]. Both assays were further integrated using the weighted nearest neighbour algorithm in Seurat[241] to create a single reduced dimensional representation of the multimodal data (Figure 3.10C).

Unsupervised clustering of this multimodal data identified cell type clusters that were relevant to unimodal projections of both gene expression and chromatin accessibility data (see Figure 3.10A-C). Heatmaps, shown in Figure 3.10D, of marker features from both assays demonstrated good purity of cell populations within each cluster. Visualisation of cells by individual and by batch also confirmed no residual batch effect (see Figure 3.10E and F). To further validate the annotation of HSPC populations I calculated the percentage of specific subpopulations of

total HSPCs for each individual and compared different phenotypes, as shown in Figure 3.11B. Although there were no statistical differences in proportions between phenotypic groups, there was a trend for HSCs being over-represented in healthy donors and a trend towards an expansion in CMPs in myeloid-biased CMML, similar to the pattern found by flow cytometry. I also performed a similar analysis of immunophenotyping data from separate vials of bone marrow mononuclear cells from the same individuals included in the single cell multiomic analysis, shown in Figure 3.11C. These data are slightly different to the multiomic analysis because the panel of antibodies used were unable to identify eosinophil/basophil/mast progenitors or megakaryocyte progenitors. However, despite these differences, this analysis demonstrated that HSCs had increased representation in all phenotypic groups in the single cell multiomic analysis, which may indicate improved detection with this technique compared to immunophenotyping, or contamination with other populations such as MPPs. Gene set enrichment analysis, using AUCell [252] to calculate an enrichment score for each cell, found high levels of enrichment for published HSC signature gene sets [253], [254] within the HSC cluster (shown in Figure 3.12B), indicating correct annotation and purity of this cluster. CMPs appeared under-represented in myeloid-biased CMML in the multiomic analysis compared to immunophenotyping, but as shown in Figure 3.11D, cluster 16 had intermediate *MPO* and *ELANE* expression and was borderline for being included in the CMP cluster. This would have changed the proportions of CMP within myeloid-biased CMML cells, because almost all the cells in this cluster were from this phenotypic group. Similarly MEP proportions in the erythroid-biased CMML group were reduced compared to the immunophenotyping data, which may be partly explained by cluster 45, which appears to have some cells that express *GATA2*, an MEP marker, but also some that express *ALAS2*, a late erythroblast marker. This cluster is almost entirely made up of cells from erythroid-biased CMML and therefore some MEP may have been included in this cluster, although the majority of cells are early erythroblasts. Therefore differences between immunophenotypic and multiomic annotation are potentially secondary to nuances of technology, although

it is also possible that malignant populations have an aberrant immunophenotype, which is frequently found in CMML[318], and the multiomic analysis may more accurately represent the cell states present in CMML patient bone marrow.

HSC subcluster analysis

The previous analysis of bulk sorted Lin^- $CD34^+$ HSPCs identified phenotypic differences between CMML patients based on the chromatin accessibility profiles of their HSPCs. Immunophenotyping analysis offered an explanation for this by identifying differences in HSPC subpopulations between phenotypes. However, an alternative explanation would be that CMML phenotypes are driven by inherent differences in HSC function, which leads to altered differentiation and accumulation of particular HSPC subpopulations. To investigate this possibility further, I conducted a subcluster analysis on HSCs. I found that HSCs from each phenotypic group clustered separately by both gene expression and chromatin accessibility. This was particularly marked in the analysis of gene expression and in both assays erythroid-biased CMML HSCs clustered more closely to healthy donor HSCs than to myeloid-biased CMML HSCs (See Figure 3.12C. I performed differential gene expression between phenotypic groups and found that erythroid-biased CMML HSCs had 214 up-regulated and 244 down-regulated genes compared to healthy donor HSCs. myeloid-biased CMML HSCs had 360 up-regulated and 375 down-regulated genes compared to healthy donor HSCs, as shown in Figure 3.12D and E. Interestingly some of the most differentially expressed genes were cell surface receptors, such as *CAMK2A* and *IL1RAP*, suggesting that CMML HSCs may experience altered interactions within the niche compared to healthy donor HSCs (Figure 3.12I and J). Gene set enrichment analysis using published MEP[313] and GMP[253] signatures confirmed that erythroid-biased CMML HSCs were enriched for MEP signature as well as GMP signature, and myeloid-biased CMML HSC were negatively enriched for MEP signature and positively enriched for GMP signature (Figure 3.12F). Gene set enrichment analysis of MSigDB Hallmark gene sets found enrichment for MYC targets, E2F targets and G2M checkpoints in

both CMML phenotypes, indicating that CMML HSCs are more proliferative than healthy donor HSCs. They were also enriched for oxidative phosphorylation and glycolysis (erythroid-biased CMML HSCs), indicating a potential metabolic switch to aerobic respiration, which would facilitate HSC proliferation and differentiation. Both CMML HSC phenotypes were also enriched for interferon alpha and gamma response, suggesting inflammatory interferon signalling may play a role in promoting this change, although interpretation is made more complex by erythroid-biased CMML HSCs being negatively enriched for inflammatory response genes and TNF-α signalling via NFκB (see Figure 3.12G and H).

Differential chromatin accessibility analysis of MACS2 called peaks found 194 more accessible peaks and 423 less accessible peaks in erythroid-biased CMML HSCs compared to healthy donor HSCs, and 3742 more accessible peaks and 2826 less accessible peaks in myeloid-biased CMML HSCs. Despite this striking difference in numbers of differentially accessible peaks, motif enrichment analysis within these peaks found C/EBP family and SPI family motifs to be highly enriched in HSCs from both CMML phenotypes, indicating that they are primed for myelomonocytic differentiation. In contrast, more accessible peaks in healthy donor HSCs were enriched for motifs of AP-1 complex members from the FOS and JUN families, where motif enrichment has previously been found in aging HSCs[254] (see Figure 3.13C and D). GREAT gene ontology enrichment analysis of peaks that were more accessible in healthy donor HSCs found terms involved in cell-cell communication, such as extrinsic apoptotic signalling pathway, positive regulation of chemotaxis and pathway-restricted SMAD protein phosphorylation were enriched, suggesting that healthy donor HSCs are primed to respond to TNF and TGF-β signalling more than CMML HSCs. Myeloid-biased CMML HSCs however had enrichment in open chromatin for pathways involved in myeloid leukocyte function and regulation of inflammatory signalling via TNF and IL-1β (Figure 3.13F. GREAT analysis for erythroid-biased CMML HSCs found no enriched terms.

3.2.7 Alterations to the gene regulatory landscape explain how different phenotypes develop within CMML

Clones that subsequently transform into CMML originate from $CD34^+$ $CD38^-$ HSC/MPPs[15]. TET2 mutations are frequently the original mutation to develop within these multipotent cells, disrupting their ability to demethylate CpG[16]. Methylated CpG are inherited by their progeny, through the activity of DNMT1 [319]. Therefore, if aberrant methylation is responsible for priming TET2m CMML HSCs for myelomonocytic differentiation, this epigenetic priming should be inherited throughout haematopoiesis and be detectable by tracking persistent differences between phenotypic groups throughout differentiation pseudotime. In order to identify evidence of such epigenetic priming I applied Slingshot to the combined weighted nearest neighbour UMAP coordinates to infer monocytic and erythroid differentiation trajectories (Figure 3.14A). Plotting cell abundance along monocyte differentiation pseudotime accurately demonstrated the expected accumulation of CMP and GMP progenitors in myeloid-biased CMML (Figure 3.14B). To identify potential drivers of phenotypic difference I combined motif enrichment analysis with gene expression of TFs in HSC subcluster. Wilcoxon test area under the curve (AUC) statistic was averaged between these assays to identify marker TFs for each phenotype (Figure 3.14C). Some of these top marker TFs, such as C/EBP-β (increased in myeloid-biased CMML) and BACH2 (increased in healthy donors), had altered motif accessibility through the first half of monocytic pseudotime (Figure 3.14D). This is suggestive of a conserved epigenetic change, originating in the HSC, altering the activity of these TFs throughout HSPC differentiation. C/EBP-β is an important activator of myeloid differentiation, particularly in the context of emergency myelopoiesis[298], whereas BACH2 is an important repressor of myeloid differentiation and promoter of lymphoid differentiation[294], [316]. C/EBP-β and BACH2 form logic gates within a gene regulatory network that are critical for determining multipotent haematopoietic progenitor cell differentiation. Both have antagonistic effects on critical myeloid genes and bind to the same sites on myeloid superenhancers. They also suppress the expression of each other, creating

a feed-forward loop supporting differentiation[295]. C/EBP-β and BACH2 also exhibit opposite binding affinity to methylated CpG within their binding motifs, with C/EBP-β preferentially binding methylated DNA and BACH2 exhibiting reduced binding, potentially linking their altered motif accessibility with TET2 loss of function[144], [146], [320]. Marker TFs identified in erythroid-biased CMML such as MEIS1 and PBX2 did not demonstrate this similar pattern, suggesting that an alternative method may be required to identify the drivers of this phenotype.

To construct a more comprehensive map of genome regulatory networks in TET2 mutant CMML I applied the Functional Inference of Gene Regulations (FigR) package to the multiomic data[257]. This package calculates correlations between gene expression and local peak accessibility to identify domains of regulatory chromatin (DORC), where peak to gene correlation is significantly different to that expected by chance. Motif enrichment is then performed to infer TF to gene links and identify whether TFs function as an activator or repressor of each target gene. This information is used to calculate a regulation score for each TF, which allows TFs to be ranked by their predicted importance to controlling gene expression within the complete data set and between individual phenotypes. Genes with more DORC associations have been shown to be enriched for genes important for coordinating differentiation[52], [257]. I therefore subset genes associated with 4 or more DORCs for each phenotype and TFs with a regulation score of >2.5 or <-2.5 to filter out TFs not involved in regulating haematopoiesis. I then performed dimensionality reduction using principal component analysis and ward clustering on TFs and genes using these principal components. Unsupervised clustering of healthy donor data identified gene and TF modules that were responsible for erythroid, myeloid, MK, T, B and NK differentiation (Figure 3.15A). Performing the same analysis on erythroid-biased CMML I found dramatic disruption to the genome regulatory landscape. Two TF modules were found that positively regulated erythroid gene expression in healthy donors, an early module containing GATA1 and TAL1, and a late module containing GATA2 and NFI family TFs. In erythroid-biased CMML however there was only one module found, made up of GATA1, TAL1 and GATA2 (Figure 3.15B).

Therefore the accumulation of MEPs in this phenotype may be caused by failure of terminal erythroid differentiation, rather than excess MEP production. B and T cell clusters were hard to identify because they were mixed with genes that were not cell type specific, suggesting weak, widespread, non-specific disruption of the normal genome regulatory landscape. The TF cluster of myeloid gene repressors, such as BACH2 and BCL11B, had diminished activity in erythroid-biased CMML, with BACH2 no longer reaching the threshold to be included on the heatmap. The cluster of genes controlling NK differentiation was expanded, which is interesting because all of these patients had increased NCAM1 expression in GM progenitors, a canonical marker of NK cells, potentially indicating that some of the NK cell differentiation programs are being utilised in myelopoiesis by TET2m CMML. I compared gene regulation scores between erythroid-biased CMML and healthy donors and found that AP-1 members JUN and JUNB, which are important for positively regulating myeloid differentiation, had increased activity in erythroid-biased CMML (Figure 3.15D). Surprisingly the activity of both C/EBP-β and C/EBP-δ were reduced compared to healthy donor networks but they remained some of the strongest activators of myeloid differentiation because of a global reduction in activity of other strong positive regulators (Figure 3.15G). In myeloid-biased CMML however there was an increase in the total number of TFs positively regulating myeloid differentiation (226 vs 176). Some important activators of myeloid differentiation, such as C/EBP-β, FOSL2 and SPI1, have increased predicted activity, and the important inhibitor of myeloid differentiation, BACH2, has decreased activity (Figure 3.15E). Erythroid regulators GATA1 and TAL1 also have reduced activity, and as a result myeloid activating TFs are the strongest regulators within the network of myeloid-biased CMML ((Figure 3.15H). These changes help to explain the clinical phenotype seen in myeloid-biased CMML, where CMP or GMP cells accumulate and myeloid leukocytes are produced in preference to cells from other lineages. The increased importance of C/EBP-β suggests myeloid-biased CMML is reliant on emergency myelopoiesis pathways during 'steady-state' haematopoiesis,

which may also explain the exaggerated neutrophil and monocyte response seen in patients with CMML when they have an inflammatory insult.

3.3 Discussion

In this chapter I sought to characterise HSPC dysfunction in TET2m CMML by performing deep molecular phenotyping with multiomic assays on primary patient samples. CMML HSPC populations were heterogeneous between patients and on this basis patients can be divided into 3 groups, erythroid-biased, myeloid-biased and balanced (normal-like)[17]. I demonstrated for the first time that these differences can be detected using distinct chromatin accessibility signatures in bulk HSPC samples. Gene ontology analysis of differentially accessible chromatin showed that HSPCs from erythroid-biased and myeloid-biased groups were primed for differentiation towards these lineages. Transcription factor motif enrichment identified GATA family motifs in open chromatin of erythroid-biased TET2m CMML samples, and C/EBP/ETS family motifs in myeloid-biased TET2m CMML samples in keeping with the known function of these TF families[174], [284].

I surveyed HSPC composition by immunophenotyping of BM-MNC from TET2m CMML and healthy donors and found that CMP populations were reduced and MEP populations increased in TET2m CMML. This is at odds with previous reports in non-genotype restricted CMML patients, where either no difference in HSPC composition, or increased CMP populations were observed[16], [211]. Both of these previous studies were conducted on a relatively small number of individuals and prior to the discovery of patient-specific HSPC phenotypes within CMML. Therefore the differences between their findings and those reported in this thesis may be explained by inter-patient heterogeneity and sample size. When I explored this heterogeneity further using correlation between subpopulation proportions in TET2m CMML I found that patients tended to have expansion of either CMP, GMP or MEP populations and that when I split patients up by groups defined by alterations in their HSPC chromatin accessibility, that myeloid-biased patients

had a trend towards CMP or GMP expansion and erythroid-biased patients had an increase in MEP populations.

Gene expression data alone was unable to identify different TET2m CMML patient HSPC phenotypes on dimensionality reduction and clustering. However gene set enrichment analysis, in individuals grouped by chromatin accessibility clusters, found enrichment for signatures of erythroid progenitor gene sets in erythroid-biased patients and neutrophil and myeloid progenitor gene sets in myeloid-biased patients. Chromatin is reorganised prior to gene expression change during differentiation, therefore these findings are in keeping with HSPC lineage priming[174]. To explore this further I interrogated HSCs by single cell multiomic analysis and found that, by both gene expression and chromatin accessibility, HSCs formed unsupervised clusters in their pre-determined HSPC phenotypic groups. Gene set enrichment analysis of gene expression data was able to detect evidence of erythroid priming in erythroid-biased CMML HSCs. Both CMML groups were enriched for myeloid (GMP) genes, indicating that the lineage priming identified in bulk HSPCs is already present at the earliest haematopoietic progenitor, the HSC, in keeping with previous work identifying the HSC as the CMML cell of origin [15], [16]. TF motif enrichment of open chromatin in both erythroid-biased and myeloid-biased CMML HSCs found C/EBP family motifs to be the most strongly enriched. Erythroid-biased CMML HSCs did not display strong enrichment for GATA family motifs despite the gene expression findings indicating erythroid priming. These analyses were conducted comparing CMML subtype with healthy donor HSCs and therefore it is possible that GATA family motifs, which were enriched in erythroid-biased CMML HSPCs in bulk analysis, were strongly enriched in both healthy donors and erythroid-biased CMML HSCs and so it was not possible to detect a difference in enrichment between these groups.

In this first multiomic single cell analysis of CMML, I created a picture of how the haematopoietic gene regulatory networks are altered in erythroid-biased and myeloid-biased TET2m CMML. I applied the FigR package to construct correlative links between gene expression and chromatin accessibility data then

inferred gene regulatory properties of TFs by performing motif enrichment analysis on linked accessible peaks. This created a comprehensive map of gene regulation, highlighting the widespread disruption of TF activity controlling erythroid terminal differentiation and the failure of myeloid suppression in erythroid-biased TET2m CMML. These findings help explain how erythroid-biased HSPCs can result in a clinical phenotype of anaemia and myelomonocytic expansion. The only previous report of erythroid-biased CMML found it was associated with less anaemia compared to other HSPC phenotypes[17]. In my own analysis this was not the case, although I have analysed fewer than half the number of CMML patients included in this study and therefore have less statistical power to detect such an association. The reduced activity of myeloid suppressors such as BACH2, requires further investigation because this would potentially create a targetable axis with which to switch off aberrant myelomonocytic differentiation. BACH2 has reduced ability to bind methylated CpG, potentially offering a link between TET2 loss of function and reduced BACH2 activity[146]. Another avenue that needs to be explored is the impact of aberrant splicing in concert with TET2 dysfunction as SRSF2 mutations were enriched in patients with erythroid-biased TET2m CMML. Potentially aberrant splicing may impact the production of disrupted TFs, or independent of TF activity, it could impact spliceasome function by altering the ability of the spliceasome complex to bind methylated CpG on pre-mRNA molecules. Further work is required to understand this more fully.

Myeloid-biased TET2m CMML on the other hand maintained normal control of erythropoiesis, but there was an expansion of TFs positively regulating myelopoiesis, alongside reduced function of myeloid suppressing TFs. Of particular interest is the increased reliance on C/EBP-β, a pioneer TF responsible for regulating emergency myelopoiesis[296], which was identified in bulk chromatin accessibility, single cell pseudotime analysis, HSC transcription factor motif enrichment and FigR gene regulatory network analysis as having increased activity in myeloid-biased CMML. Multiple studies have found that C/EBP-β binds more avidly to motifs containing methylated CpG, raising the possibility that TET2 loss of function may facilitate

the switch from C/EBP-α to C/EBP-β as the master regulator of myelopoiesis[144], [146], [320]. In keeping with this hypothesis are the findings that methylation is increased in enhancer regions of myeloid specific genes in TET2 mutant CHIP/CCUS and CMML patient derived TET2 mutant iPSC models[180], [302], [321]. Many of these same regions are also associated with open chromatin in innate immune memory in murine HSCs, facilitating the rapid deployment of C/EBP-β dependent emergency myelopoiesis in response to secondary infection[298].

Emergency myelopoiesis requires an inflammatory trigger, such as the detection of a pathogen (PAMP) or tissue damage (DAMP), or cytokines produced by the resultant immune activation[322]. Low level inflammation is also found as part of a physiological response to ageing[323]. Although the effect of this on haematopoiesis is still being fully characterised. Aged HSCs possess reduced self-renewal potential and enhanced myeloid differentiation bias compared to younger counterparts, similar to HSCs that have been chronically exposed to IL-1β[324], [325]. A potential driver of this aged phenotype is the activation of emergency myelopoiesis during steady-state periods, which provides a proliferative advantage to myleoid-biased progenitors and can promote stem cell exhaustion[290]. Such conditions are particularly advantageous to TET2 mutant clones, where loss of TET2 function replenishes HSC self-renewal potential and they gain a further proliferative advantage in response to inflammation[41]. Population wide studies of TET2 mutant CHIP also support this premise, where TET2 mutations are found particularly at an older age than other CHIP clones and their clonal expansion accelerates with age[22], [25].

Another potent driver of inflammation in CMML is the expansion of myelomonocytic cells, in particular the monocytes, which have been demonstrated to be transcriptomically pro-inflammatory in some studies and contribute to the increased serum concentrations of many inflammatory cytokines in patients with CMML[109], [110]. As I outlined in our original hypothesis, we suspect that monocytes provide a supportive environment for HSPCs in TET2m CMML. The finding that HSPCs are primed for emergency myelopoiesis is completely in keeping with this hypothesis

and the characterisation of these interactions is the focus of my next chapter. Furthermore, TET2 is a negative regulator of inflammation through its association with HDAC2[162]. My findings raise the possibility that TET2 may also be required for the resolution of emergency myelopoiesis by demethylating C/EBP-β binding motifs, thereby reducing the avidity of C/EBP-β binding and facilitating its replacement by BACH2, which in turn switches off myeloid specific gene expression.

3.3.1 Future work

Linking methylation change with gene regulatory network dysfunction would for the first time provide insight into the mechanistic role of TET2 dysfunction in CMML. In particular I am interested in altered methylation at BACH2/C/EBP-β motifs to detect whether methylation at these loci may be guiding altered gene regulatory function in TET2m CMML. This may also have wider implications on the control of emergency myelopoiesis, which is of great interest to those working on haematopoiesis and infection. I have created TET2 knockdown cell lines and have primary patient material that I will assay by whole genome methylation analysis using TET-assisted pyrimidine borane sequencing with my second supervisor and collaborator Skirmantas Kriaucionis. I will perform a correlation analysis of chromatin accessibility, using the paired samples presented in this chapter, with methylation at BACH2 and C/EBP-β binding motifs to identify the role of methylation in determining the altered activity found of these TFs in TET2m CMML.

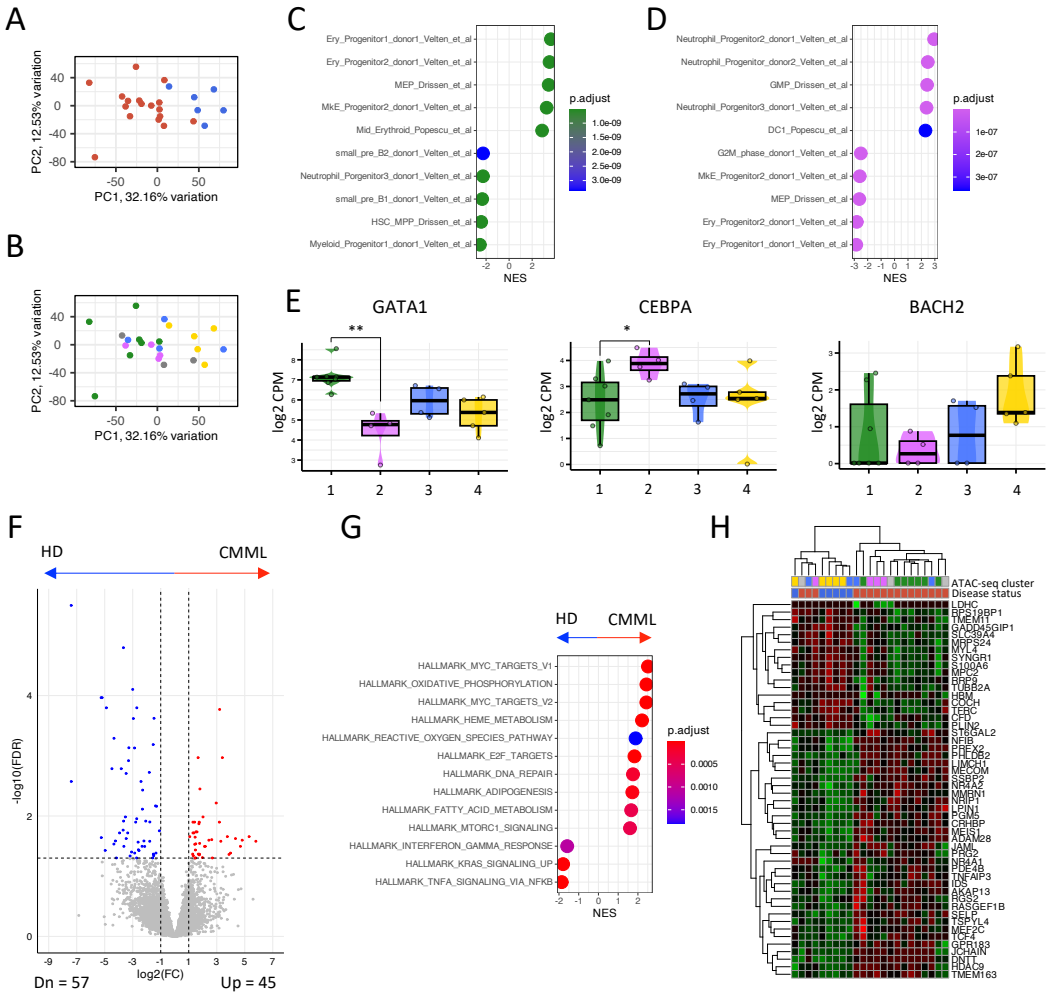


Figure 3.6: RNA-seq analysis of HSPCs supports phenotypic division of TET2m CMML HSPCs discovered by ATAC-seq but variance between samples is more strongly driven by differences with healthy donors - PCA analysis of limma batch corrected log2 CPM coloured by disease status, TET2m CMML in red and healthy donors in blue(A), and ATAC-seq clusters(B). Clusters 1-4 in green, pink, blue and yellow respectively, samples without ATAC-seq data are displayed in grey. Gene set enrichment analysis using curated gene sets of enriched haematopoietic populations comparing cluster 1 samples (C) and cluster 2 samples (D) with all other clusters. (E) Violin plots of log2 CPM normalised gene expression of TFs that had motifs highly enriched within open chromatin in cluster 1 (GATA1), cluster 2 (CEBPA) and cluster 4 (BACH2) grouped by ATAC-seq cluster. Groups compared using Wilcoxon test and only significant differences ($p<0.05$) have been highlighted with brackets. (F) Volcano plot of DESeq2 differential gene expression analysis between TET2m CMML and healthy donor HSPCs. Significant genes are highlighted in red (up) and blue (down). (G) Gene set enrichment analysis using Hallmark MSigDB curated gene sets on a gene list ranked by DESeq2 likelihood ratio between TET2m CMML and healthy donor HSPCs. (H) Top 50 differentially expressed genes between TET2m CMML and healthy donor HSPCs displayed in a heatmap. ATAC-seq cluster and disease status are annotated with the same colours described above.

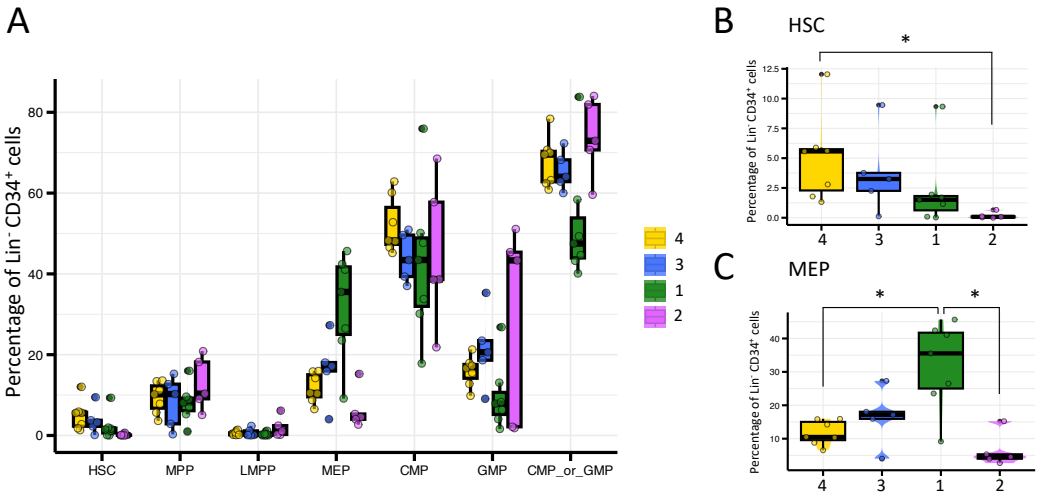


Figure 3.7: Re-analysis of HSPC immunophenotype confirms phenotypic clusters identified by chromatin accessibility - Flow cytometry data from cells sorted for ATAC-seq and RNA-seq analysis. Box and whisker plot of HSPC sub-populations as a proportion of total Lin$^-$ CD34$^+$ HSPCs (A). ATAC-seq clusters are coloured from 1 to 4 in green, pink, blue and yellow respectively. Groups were compared by Wilcoxon test and significant differences ($p<0.05$) are highlighted in HSC (B) and MEP (C) subpopulations.

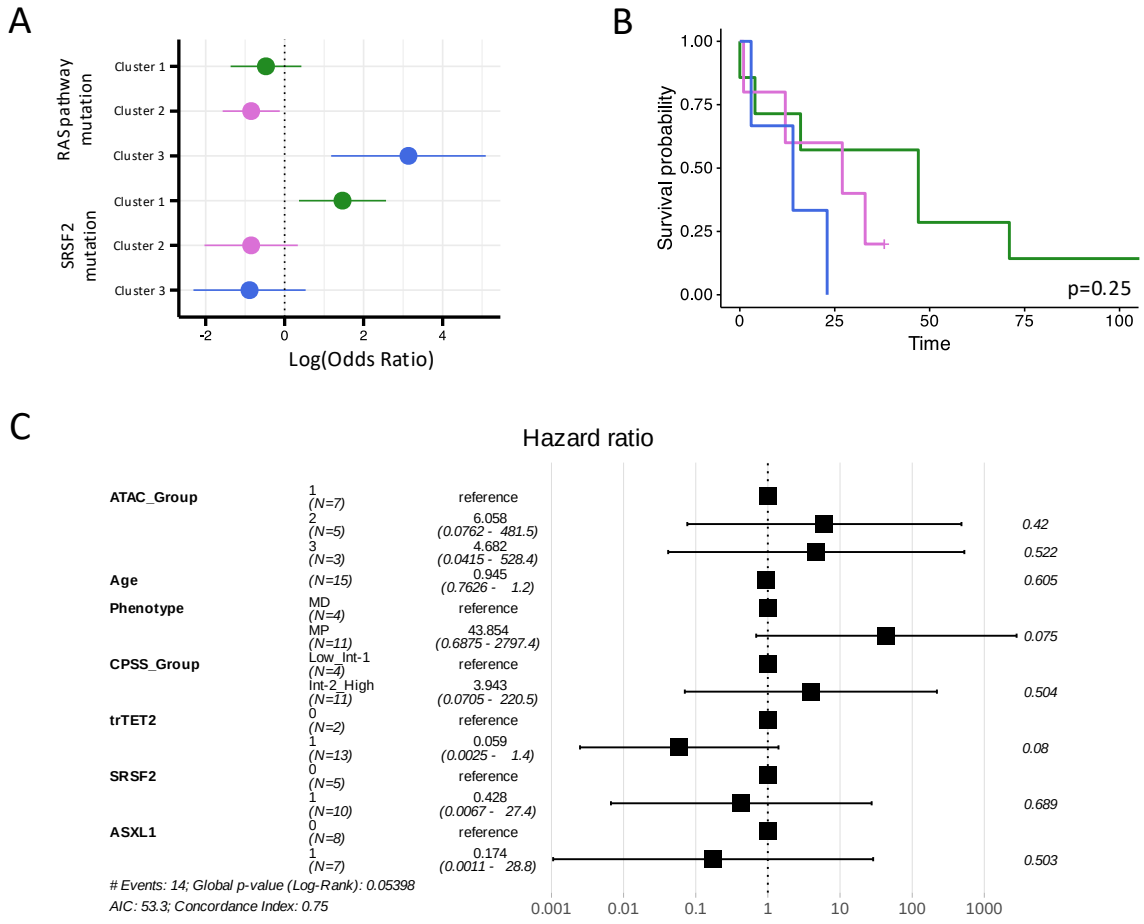


Figure 3.8: Phenotypic clusters are associated with specific genotypes - (A) Forest plot showing generalised odds ratio of RAS pathway mutation (KRAS, NRAS or CBL) or SRSF2 mutation occurring in patients within each phenotypic cluster compared to all other clusters. (B) Kaplan-Meier analysis of survival displaying 3 phenotypic clusters, cluster 1 to 3 in green, pink and blue respectively. (C) Forest plot of multivariable Cox proportional hazards regression analysis.

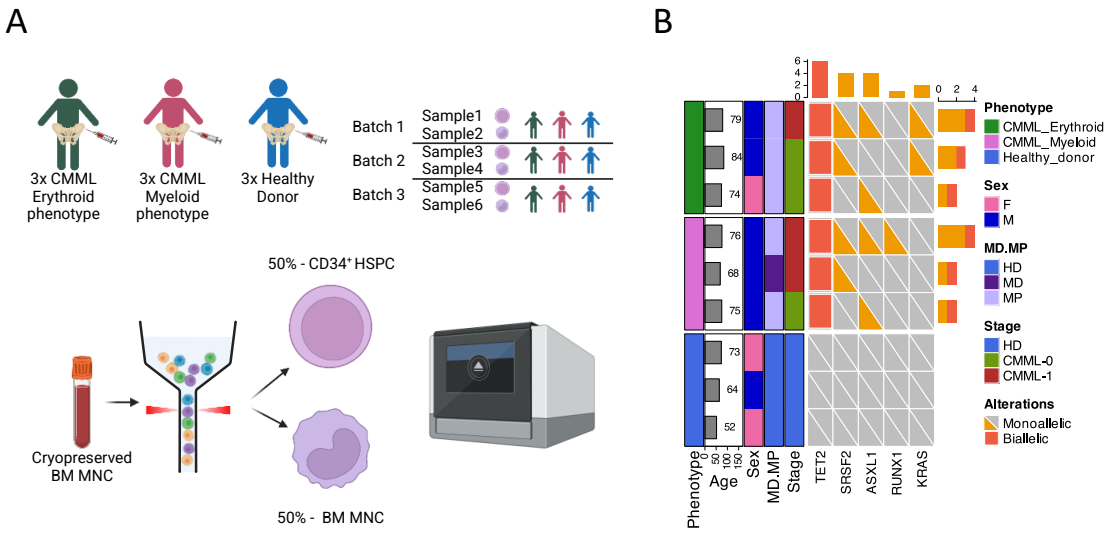


Figure 3.9: Overview of snRNA-seq and scATAC-seq multiomic profiling of TET2m CMML - (A) Summary of experiment to explore drivers of HSPC phenotypic differences in TET2m CMML (B) Demographics and mutational burden of individuals included in the experiment

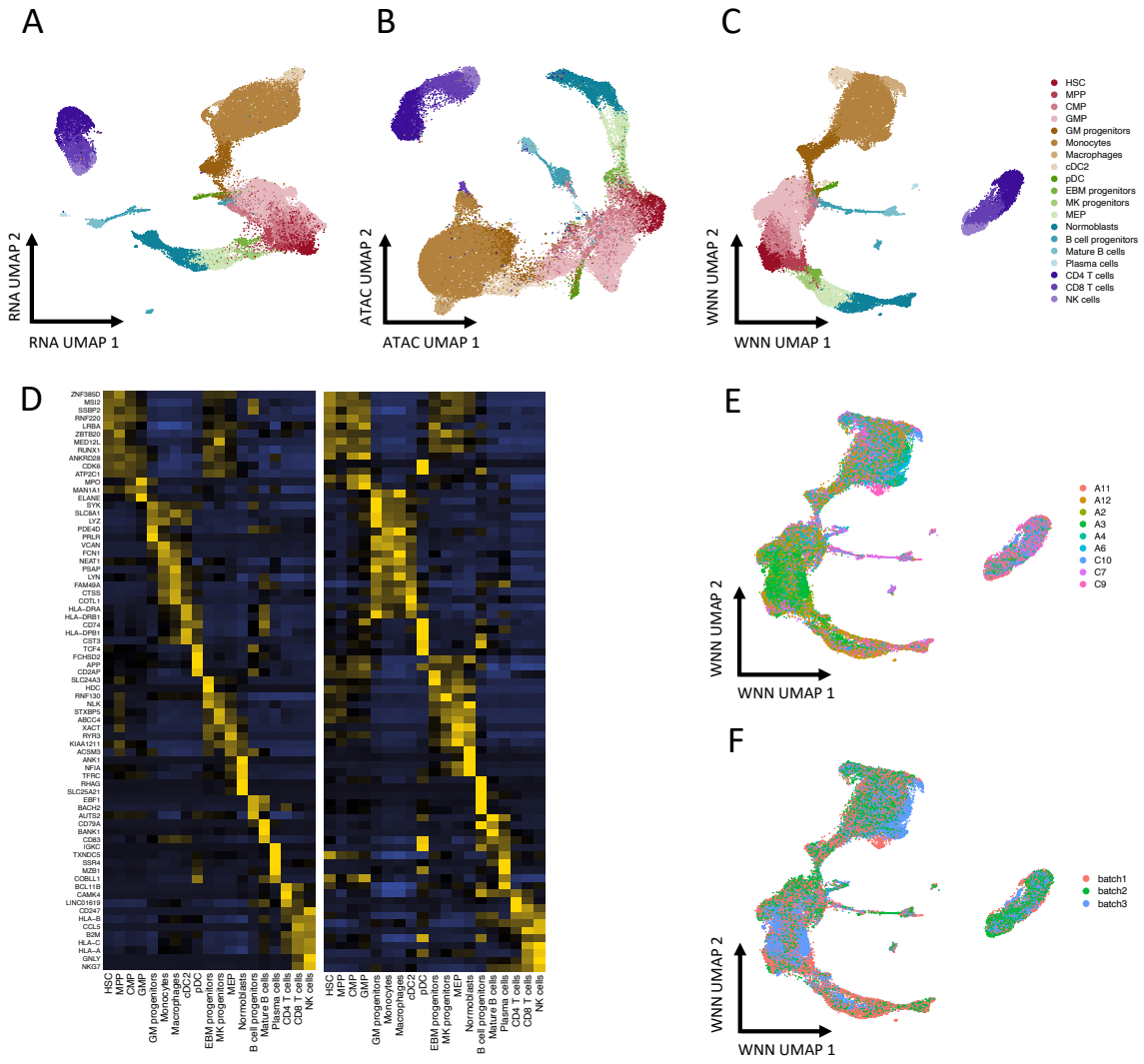


Figure 3.10: snRNA-seq and scATAC-seq integration identifies multiomically defined cell states - snRNA-seq and scATAC-seq data were normalised, corrected for inter-individual differences using Harmony and integrated using the weighted nearest neighbour algorithm. Clusters were annotated and plotted on harmony corrected RNA-seq (A), ATAC-seq (B) and combined (C) UMAP projections. (D) Heatmap of scaled marker genes expression/contig accessibility for RNA (left) and chromatin accessibility (right). WNN integrated UMAP coloured by individual (E) and batch (F) demonstrating successful integration of all samples.

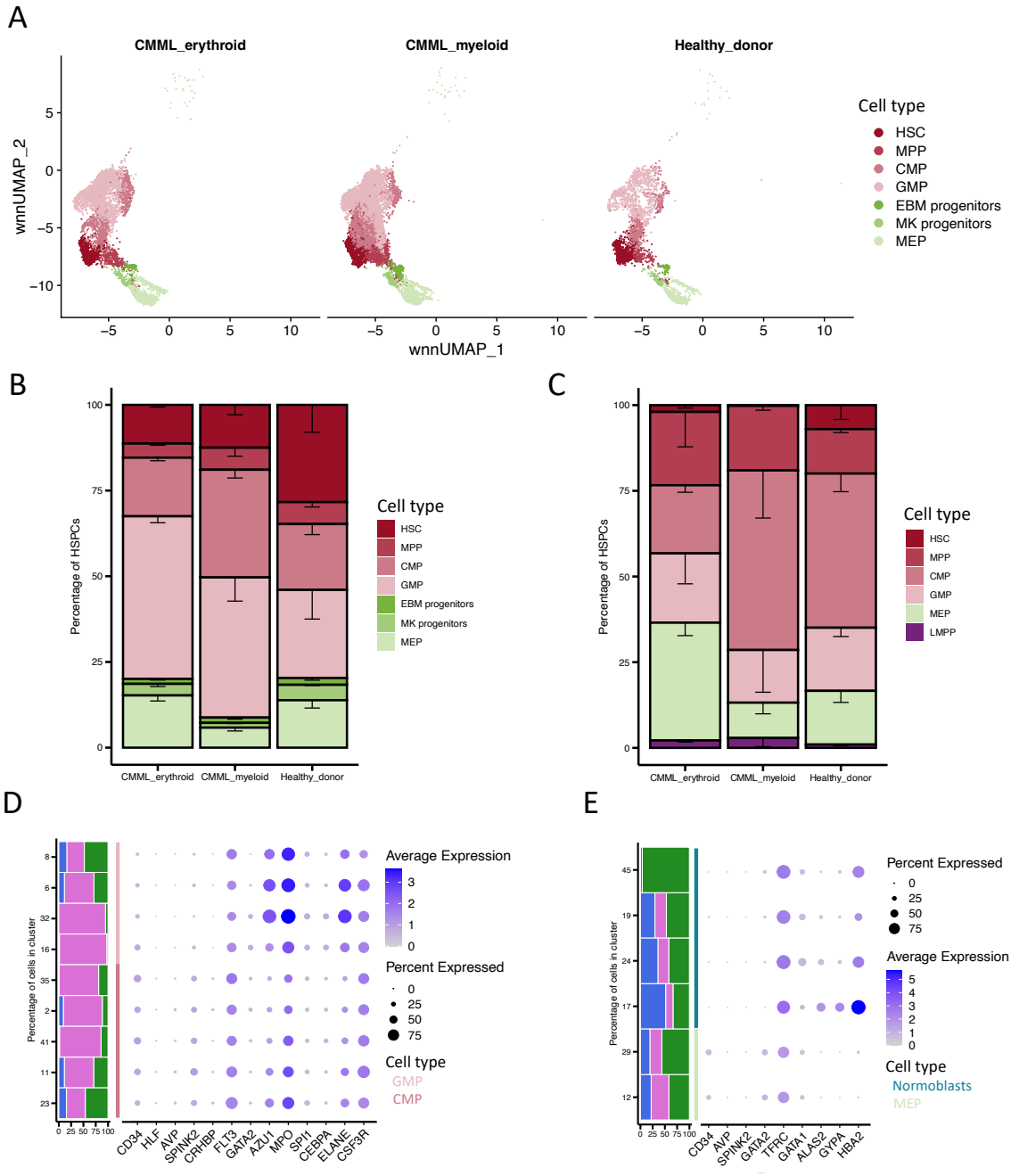


Figure 3.11: HSPC subpopulations are well represented in all phenotypic groups - (A) UMAP projection of HSPC subset of cells, split by phenotypic group. (B) Percentage of multiomically defined HSPCs per individual, split by phenotypic group. Error bar represents standard error of means. (C) Percentage of immunophenotypically defined HSPCs from Lineage$^-$ CD34$^+$ population. Data from the same individual but different vial to that used in multiomic assessment. Error bars represent standard error of means. Dotplot of marker gene expression and proportion of cells within each cluster from different phenotypic groups in stacked bar. Healthy donors, myeloid-biased CMML and erythroid-biased CMML in blue, pink and green respectively. Clusters shown were merged to make CMP and GMP clusters (D) or MEP and normoblasts (E).

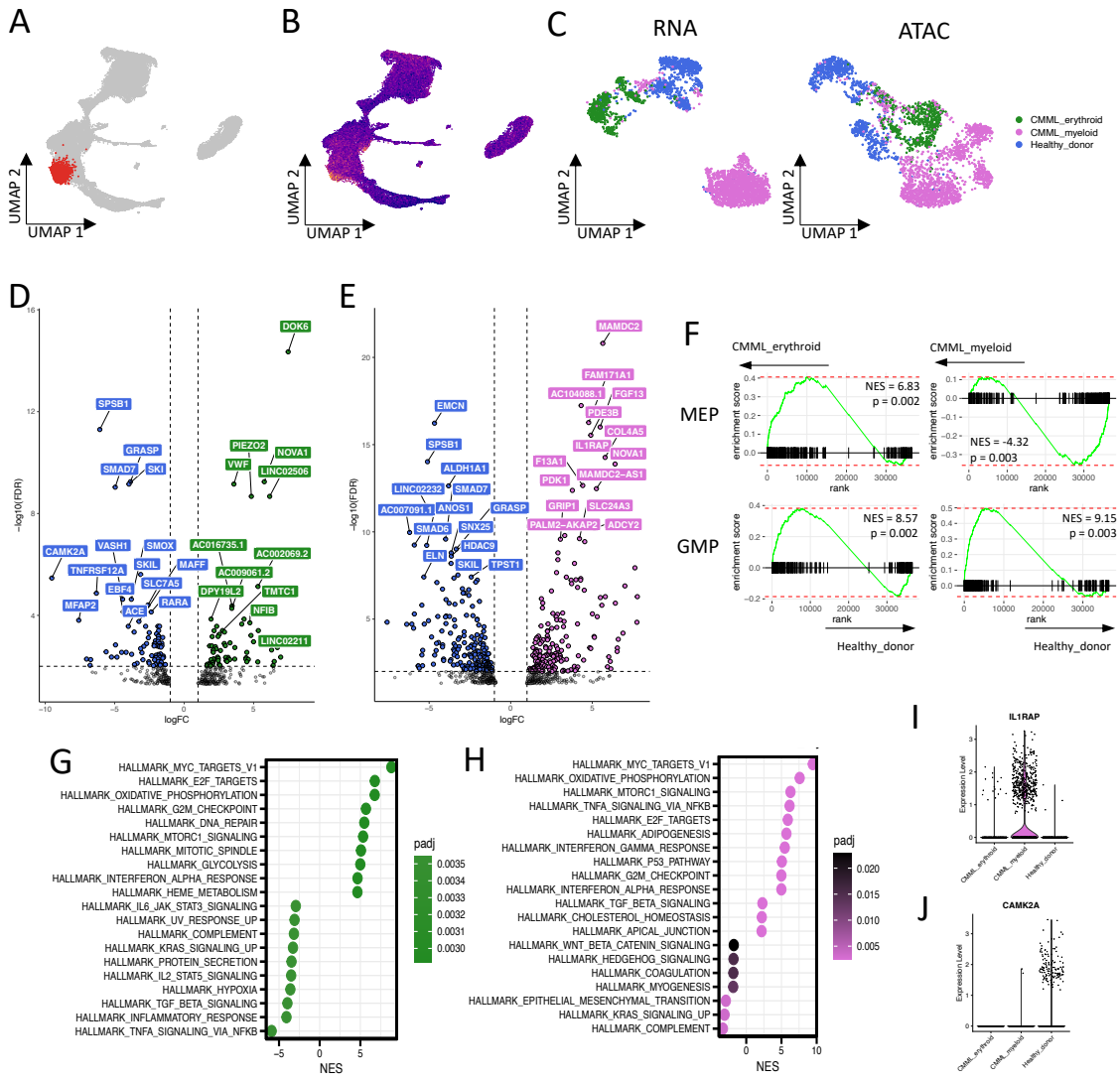


Figure 3.12: Gene expression analysis of HSC subcluster - (A) HSC cluster highlighted in red on WNN UMAP. (B) AUCell gene set enrichment of HSC vs progenitor population gene signature. Yellow is highly enriched and blue poorly enriched. (C) HSC subcluster UMAP projections of reintegrated RNA (left) and chromatin accessibility (right) data following Harmony correction for batch. Differential gene expression analysis for erythroid-biased CMML HSCs vs healthy donor HSCs (D) and myeloid-biased CMML HSCs vs healthy donor HSCs (E). (F) Gene set enrichment analysis for MEP (top) and GMP (bottom) signature genes in erythroid-biased CMML HSCs (left) and myeloid-biased CMML HSCs (right) compared to healthy donor HSCs. Gene set enrichment analysis of MSigDB Hallmark gene sets in erythroid-biased CMML HSCs vs healthy donor HSCs (G) and myeloid-biased CMML HSCs vs healthy donor HSCs (H). Violin plots of RNA expression split by phenotypic group for *IL1RAP* (I) and *CAMK2A* (J).

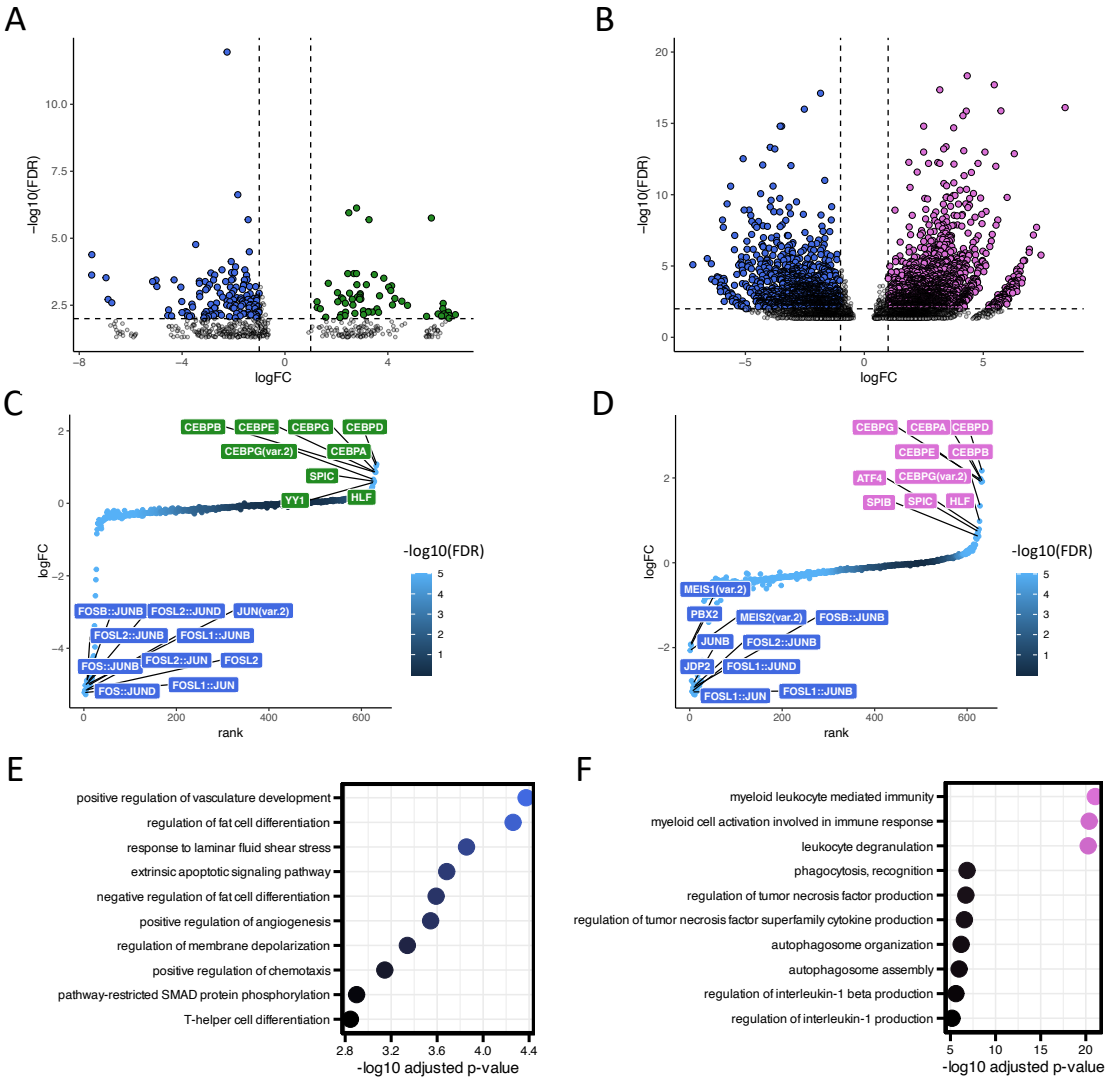


Figure 3.13: Chromatin accessibility analysis of HSC subcluster - Differential chromatin accessibility between erythroid-biased CMML HSCs and healthy donor HSCs (A) and myeloid-biased CMML HSCs and healthy donor HSCs (B). Transcription factor motif enrichment analysis within open chromatin in erythroid-biased CMML HSCs vs healthy donor HSCs (C) and myeloid-biased CMML HSCs vs healthy donor HSCs (D). GREAT gene ontology analysis of differentially open chromatin in healthy donor HSCs (E) and myeloid-biased CMML HSCs (F).

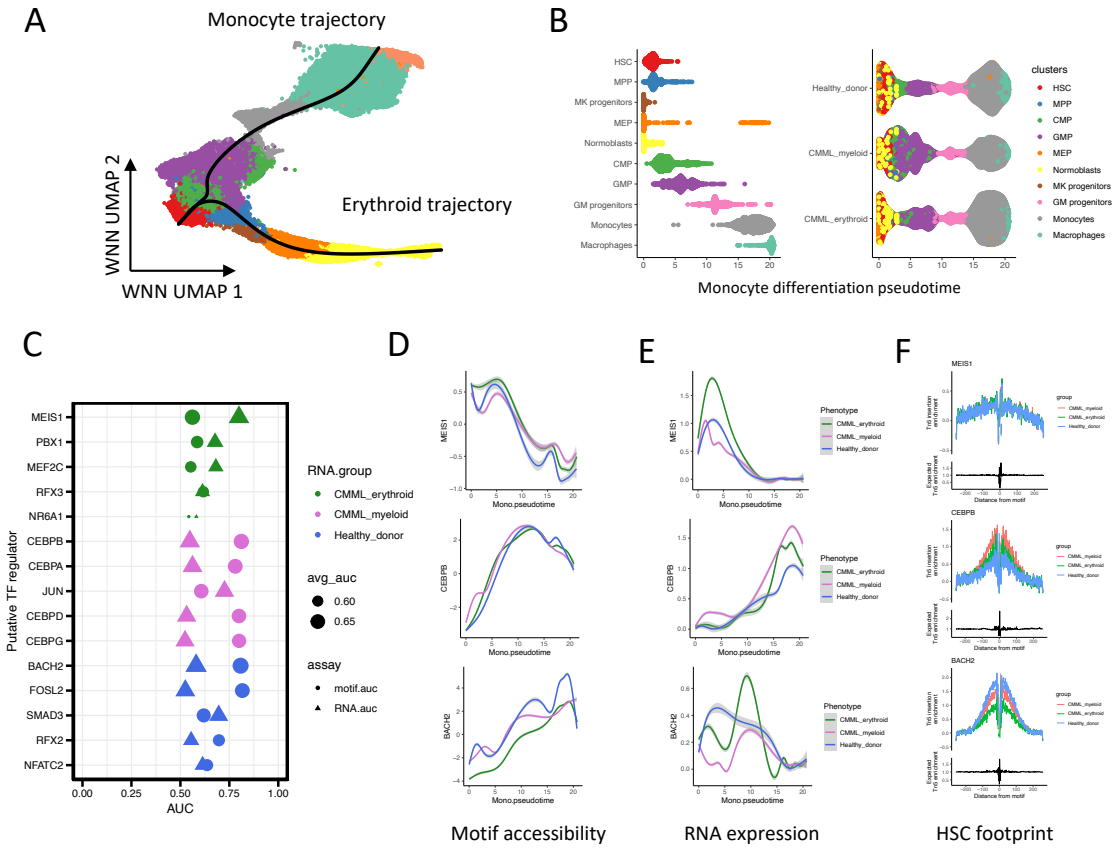


Figure 3.14: Exploring potential drivers of altered HSC function in CMML - (A) Slinghshot identified differentiation trajectories on weighted nearest neighbour UMAP. (B) Position of individual cells along myeloid differentiation pseudotime by cell type (left) and phenotype (right). (C) Top 5 marker TFs for HSCs in each phenotype identified by Wilcoxon rank test between CMML phenotypes and healthy donors for TF motif enrichment of accessible peaks and TF RNA expression. Average area under the curve statistic from both tests was used to rank TF markers for each phenotype. Deviation score for TF motif enrichment in accessible peaks (D) and normalised TF RNA expression (E) of top marker TF plotted through monocyte differentiation pseudotime for each phenotype. (F) TF motif footprint for top TF marker in each phenotype.

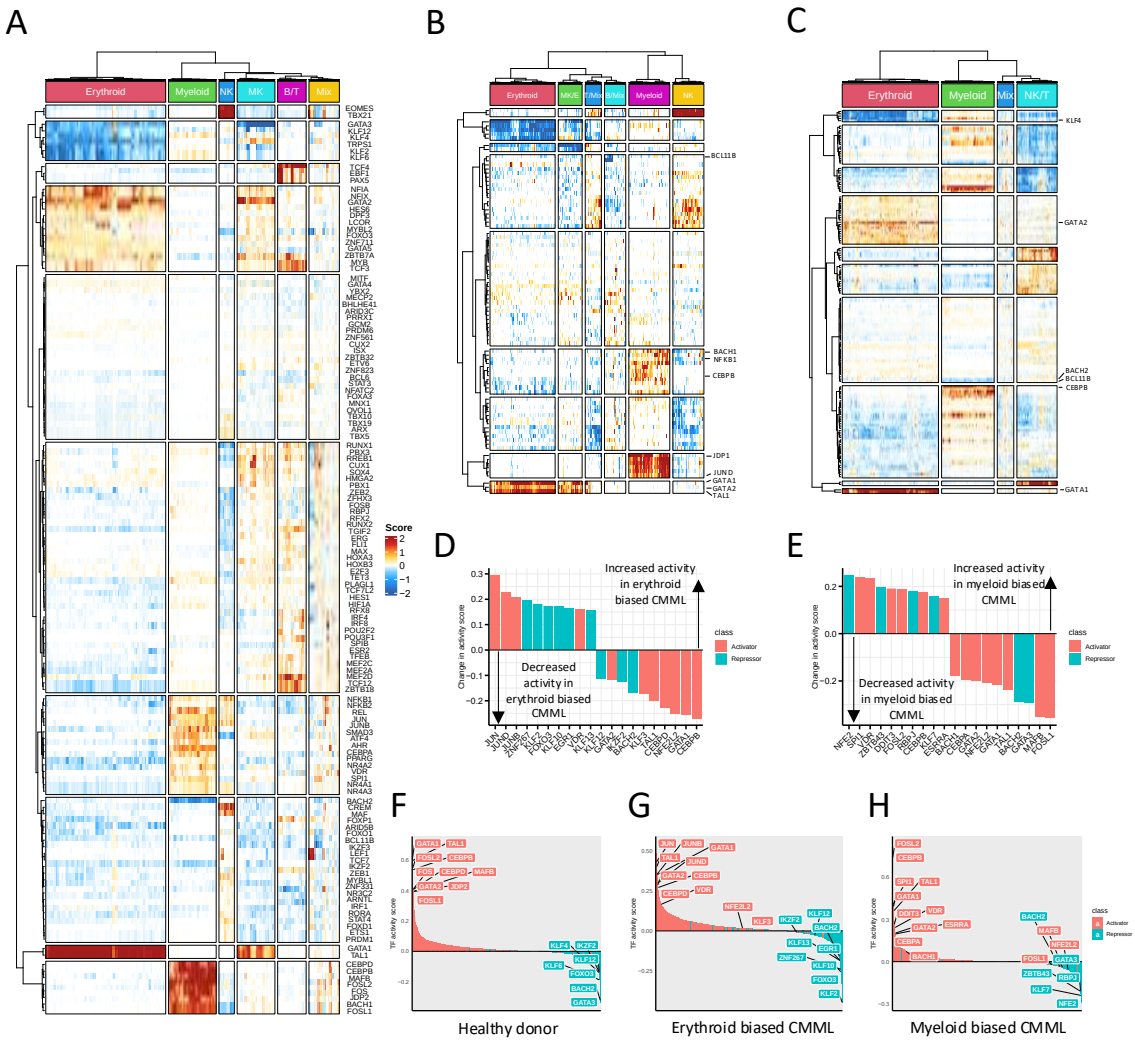


Figure 3.15: Erythroid and myeloid-biased CMML are driven by distinct changes to the genomic regulation landscape - Heatmaps of TF by DORC genes coloured by scaled regulatory score in healthy donor cells (A), erythroid-biased CMML (B) and myeloid-biased CMML (C). Heatmaps are split into TF and gene modules by ward clustering on principal components and annotated by overrepresentation analysis using cell type gene signatures. Top 10 TFs with increased and decreased activity comparing erythroid-biased CMML (D) and myeloid-biased CMML (E) with healthy donors. TFs are coloured by class, with activators being TFs with positive average regulation score across all phenotypes and repressors with negative average regulation score. TF regulation scores were ranked from highest to lowest and waterfall plots were constructed for healthy donors (F), erythroid-biased CMML (G) and myeloid-biased CMML (H).

4

Characterising changes to the extrinsic regulation of haematopoiesis in TET2m CMML

Contents

4.1 Introduction 116
4.1.1 Extrinsic regulators of haematopoiesis 116
4.1.2 Factors driving emergency myelopoiesis 117
4.1.3 Structural changes to bone marrow architecture in CMML 118
4.1.4 Non-haematopoietic stromal cells display altered abundance and function in CMML 119
4.1.5 Altered cytokine signalling in CMML 120
4.1.6 Mature myeloid cell populations undergo phenotypic changes in CMML 122
4.1.7 Summary and plan for investigation 124
4.1.8 Aims 124
4.2 Results 125
4.2.1 TET2 mutant CMML monocytes are transcriptionally and epigenetically distinct from healthy donor monocytes 125
4.2.2 snRNA-seq facilitates identification of putative aberrant cell-cell interactions in TET2m CMML bone marrow . . 130
4.2.3 Classical monocyte populations are expanded 10 fold in TET2m CMML bone marrow mononuclear cells and exhibit an aberrant immunophenotype 132
4.2.4 TET2m CMML is associated with an increase in both classical monocyte and macrophage populations in the bone marrow 135
4.2.5 HSPCs reside in close proximity to classical monocytes and macrophages in TET2m CMML bone marrow . . . 140

4.2.6 Multinichenet analysis identifies putative interactions that are predicted to promote HSPC dysfunction in TET2m CMML 140
4.3 Discussion 150
4.3.1 Conclusion 156

4.1 Introduction

4.1.1 Extrinsic regulators of haematopoiesis

Steady-state haematopoieis occurs almost exclusively within the bone marrow of adults[326]. Haematopoietic cells reside intermixed with stromal cell populations in a highly complex and dynamic structure. Interactions between different cell types are inherently important for guiding haematopoietic cell migration, proliferation and differentiation. This phenomenon can be readily demonstrated through *in vitro* colony forming assays, where HSPCs are seeded into a semi-solid methylcellulose culture supplemented by cytokines[269]. In these cultures haematopoietic progenitors can be supported to both expand and differentiate into mature populations, with certain combinations of cytokine facilitating the proliferation and differentiation of progenitors into specific mature haematopoietic populations. Colony forming assays however fail to sustain indefinite haematopoiesis and therefore do not provide the support required to maintain HSC self-renewal. However, these conditions can be re-created in long term cobblestone area forming cell (CAFC) assays, by seeding HSPCs on to a feeder layer of irradiated stromal cells. These stromal cells support HSC growth and self-renewal through membrane bound cell-cell interactions, such as via NOTCH ligands and CXCL12, and by paracrine release of supportive cytokines such as Stem Cell Factor (SCF)[327]–[329]. More recent efforts to optimise conditions for *ex vivo* HSC expansion have identified the importance of several ligand-receptor interactions such as WNT[330] and prostaglandin[331] signalling as well as mechanoreceptor interactions with polymer scaffolds for maintaining HSC expansion and self renewal[332]. Haematopoiesis therefore relies on the highly

complex and coordinated interplay between HSPCs and their surrounding cells and structures to maintain normal function.

Recent advances in imaging technologies have facilitated the unsupervised identification of cellular neighbourhoods that exist within human bone marrow and thereby the topological associations between the different branches of haematopoiesis[333]. The location of HSCs remains controversial, although they are increasingly accepted as residing within a peri-sinusoidal hypoxic niche, in keeping with their reliance on anaerobic metabolism to maintain quiescence[334]. Multiple lines of evidence have linked HSCs to residing in close proximity to endothelial cells, $CXCL12^{+}$ abundant reticular cells, $NESTIN^{+}$ mesenchymal stromal cells, neurons and macrophages, each of which perform a role in maintaining stem cell function[335]. Cycling HSCs switch metabolism to aerobic glycolysis and migrate to more oxygen rich areas located between endosteal and sinusoidal niches that are populated by a mixture of all other haematopoietic cell types and adipocytes. GMP and MEP relocate again to the most oxygen rich endosteal/arteriolar niche, where they facilitate the formation of clusters of MPO^{+} myeloid progenitors and immature erythroid blood islands respectively. Terminal differentiation of all lineages tends to occur close to sinusoids, with some separation created by different sinusoids providing distinct supportive factors, such as M-CSF[336], [337].

4.1.2 Factors driving emergency myelopoiesis

As introduced in the previous chapter, EM can be triggered by a combination of damage/pathogen response signals and cytokines resulting in a characteristic 'left-shifted' picture with myeloid precursors being released into the peripheral blood alongside greater numbers of neutrophils and monocytes. Many models of microbial infection and sterile inflammation have been developed and multiple factors have been identified, such as G-CSF, M-CSF, IL-1 and interferon, as integral to supporting EM[338]–[341]. However, the impact of external stressors on the bone marrow architecture has only been characterised quite recently. *Wu et al* used either G-CSF or *Listeria monocytogenes* infection to provoke EM[337]. They

used live imaging of murine sternal bone marrow *in vivo* to track topographical changes to haematopoiesis and found that overall, the bone marrow structure and positioning of HSPCs remained similar to steady state. Multipotent and oligopotent HSPC remained evenly dispersed throughout the bone marrow, and lineage-specific production sites being segregated to different sinusoids. However, there was a dramatic increase in the number of granulocytic progenitors in response to G-CSF and granulocytic, monocytic and dendritic precursors after infection. Myeloid primed multipotent progenitors were observed to form small clusters in response to infection, but this was not seen in other conditions. Infection also resulted in an increase in the number of mono/DC and granulocytic production sites and reduced number and size of lymphoid and erythroid production sites. G-CSF resulted in the reduced number and size of production sites of all lineages. These changes all rapidly reversed following the resolution of the inflammatory challenge, demonstrating the dynamic nature of haematopoiesis to rapidly adapt to the specific demand created by an insult.

4.1.3 Structural changes to bone marrow architecture in CMML

Cytological and histological assessment of bone marrow aspirate and trephine are fundamental to making a diagnosis of CMML. Therefore many descriptive studies exist of qualitative and quantitative changes to morphologically or immunohistochemically defined haematopoietic populations[342]. However, there are a paucity of data comparing the spatial relationships between these populations in CMML with healthy donors. Bone marrow cellularity is almost always increased in CMML, with a concomitant reduction in adipocytes. $CD34^+$ HSPCs can be increased and occasionally form clusters, which can lead to disparity between cytological and histological bone marrow blast quantification. Blasts, morphologically defined immature haematopoietic progenitors, are usually $CD34^+$ in myeloid malignancies and their quantitation is an important predictor of disease prognosis. However, sometimes blasts are $CD34^-$ $CD117^+$ in CMML, indicating a more mature phenotype

that probably overlaps with pro-monocytes, which can be morphologically difficult to distinguish from blasts without the addition of cytochemical stains such as naphtyl butyrate esterase[343]. Monocytic precursors, Monocytes and macrophages are increased in CMML. Neutrophils and granulocytic precursors are also increased, with evidence of dysgranulopoiesis such as nuclear hypolobation and hypogranularity. Megakaryocyte numbers are usually increased, sometimes with evidence of clustering and emperipolesis. They can also exhibit dysplasia, with reduction in size and reduced nuclear lobation[342]. Clusters of CD123 expressing plasmacytoid dendritic cells are found in around 10% of CMML patients and clusters of spindle shaped $CD25^+$ $CD117^+$ mast cells can also be found in SM-CMML[343], [344]. A detailed assessment of topological changes within CMML bone marrow has not yet been formally undertaken, in particular it is not known whether HSPCs reside within an altered niche in CMML.

4.1.4 Non-haematopoietic stromal cells display altered abundance and function in CMML

Non-haematopoietic stromal cell populations are quantitatively and functionally altered in CMML. Mesenchymal stem cells (MSC) have been found to be present at a reduced frequency in CMML and be functionally senescent compared to healthy donor counterparts, with reduced capacity for proliferation and decreased production of haematopoietic stem cell supporting factors such as CXCL12, ANGPT1 and FRZB[345]. However, CMML MSCs also overexpress some potential pathological factors, such as hepatocyte growth factor (HGF) and hypoxia-inducible factor 1-alpha (HIF1A), that have been implicated in promoting the development of AML and MDS respectively[346], [347]. Furthermore, co-culture of CMML MSCs with donor cord blood HSPCs skews their differentiation towards myeloid-lineage and reduces erythroid colony formation, phenocopying changes to colony forming assays using primary CMML patient HSPCs[114], [348]. Cross talk between MSCs and HSPCs has also been implicated in MDS and JMML pathogenesis through both direct and indirect mechanisms. MSCs in MDS directly influence HSPC function

via reduced production of HSPC supportive cytokines, such as CXCL12, IGFBP2, IGF1, SCF and TPO[349]. Whereas MSCs in a murine model of *Ptn11* mutant JMML exert an indirect effect on HSPC function by producing CCL3, which recruits monocytes to the HSC niche, that in turn release IL-1β that activates HSCs and promotes their differentiation[350].

4.1.5 Altered cytokine signalling in CMML

Cytokine levels in blood and bone marrow are altered in CMML. In particular pro-inflammatory cytokines - IL-8, IL-6 and TNF-α; chemokines - CXCL10 and CCL2; and growth factors - hepatocyte growth factor (HGF), macrophage colony stimulating factor (M-CSF) and vascular endothelial growth factor (VEGF) are increased[109]. Many of these factors are produced by haematopoietic cells and create a nurturing environment for HSPC growth. This creates an environment where CMML peripheral blood mononuclear cells, unlike healthy donor PBMNCs, can form spontaneous colonies in semi-solid culture without the addition of exogenous cytokines[351]. Detailed analysis of the secreted factors important for supporting this process have identified IL-6, IL-8 and GM-CSF as enhancers of colony formation and IL-10 as suppressive[109], [114], [352]. The inhibitory effect of IL-10 on colony formation appears to be mediated by its suppression of GM-CSF expression, and colony formation can be restored by the addition of exogenous GM-CSF[353]. Furthermore, cytokine levels have been demonstrated to have clinical importance, where reduced IL-10 concentration in peripheral blood plasma is associated with inferior survival and the addition of IL-10 concentration to the Mayo CMML prognostic model improves its performance[109].

Cytokines are therefore inherently important for CMML pathobiology and identifying the cells responsible for cytokine production may improve treatment strategies. Removal of either monocytes or T cells from PBMNCs almost completely abrogates the ability of CMML patients to form spontaneous colonies in semi-solid culture, suggesting these are the cells most important for secreting supportive

cytokines[353]. RNA sequencing of CMML monocytes has found increased expression of pro-inflammatory cytokines *IL-1A*, *IL-1B*, *IL-6*, *CXCL8* (IL-8), *CXCL10* (IP-10) and *VEGFA* compared to healthy donor monocytes, suggesting monocytes are the source of many of the cytokines found to be elevated in CMML patient serum. This validates the normalisation of monocyte counts as a therapeutic goal in the IWG CMML response criteria[109], [110].

HSPCs also alter their cytokine receptor expression in CMML, enhancing their ability to respond to changes in cytokine gradients in the bone marrow microenvironment. Screening of RNA sequencing data has found 51 receptors with altered expression in CMML, including CSF1R (M-CSFR), CSF2RB (GM-CSFR), CSF3R (G-CSFR), CCR2, CXCR2, IL-15RA and TNFRSF1B, which were also confirmed to be more abundant on CMML HSPC cell surfaces compared to healthy donor controls[17]. In a separate study focussed on $CD34^+$ $CD38^-$ HSC/MPPs, known to include the leukaemic stem cell population in CMML, IL-1 receptor accessory protein (IL-1RAP) and CXCR4 were upregulated in CMML. However there was no evidence of change to M-CSFR, GM-CSFR or G-CSFR expression, highlighting the heterogeneity in cell surface receptor expression between subgroups of HSPC[15]. CXCR1 and CXCR2 are downregulated by CMML HSPCs making them more resistant to CXCL8 signalling, which inhibits colony formation in healthy controls[111]. Given these findings, it is likely that extrinsic regulation of HSPC function by monocytes is critical for CMML pathogenesis. Many cytokine signalling pathways have been implicated, but the current evidence is fragmented, focusing either on changes to HSPC function or monocyte function rather than a coordinated approach. Many of these studies have highlighted the inter-individual heterogeneity, and therefore the best way to truly understand how these interactions impact CMML pathogenesis is to study all the relevant populations within the same individual.

4.1.6 Mature myeloid cell populations undergo phenotypic changes in CMML

Monocytes are mononuclear phagocytic cells produced by differentiation of GMPs or MDPs (Monocyte-Dendritic cell progenitors) in the bone marrow. The great majority of monocytes produced are $CD14^{+}$ $CD16^{-}$ classical monocytes, which egress from the bone marrow upon maturation[354]. Classical monocytes are innate immune effector cells that can phagocytose pathogen and detect inflammation through CCL2-CCR2 chemokine signalling that triggers them to migrate out of the vasculature into tissues. This process is also important in the absence of inflammation for replenishing tissue-resident macrophage populations. Furthermore, classical monocytes can differentiate into $CD16^{+}$ $CD14^{-}$ atypical monocytes that are long-lived blood resident macrophages that patrol the vascular endothelium[355]. Classical monocytes are therefore highly plastic, non-terminally differentiated cells that adapt their function in response to micro-environmental cues. Macrophages on the other hand are tissue resident phagocytic cells that are highly specialised to their particular niche, performing homeostatic roles in steady-state and coordinate both the immune response and tissue repair following local insult. Macrophages of most tissues consist of a mixture of yolk sack derived macrophages, which form a life-long pool of self-replenishing macrophages, and monocyte derived macrophages. The ratio of yolk-sack derived and monocyte derived macrophages are different for each organ. In the bone marrow, macrophages are almost exclusively monocyte derived and specialise into multiple niche specific subtypes. The best described role of macrophages within the bone marrow is their role as nurse cells, coordinating maturation of normoblasts to erythrocytes within erythroid blood islands[356]. However, macrophages are also able to directly regulate HSC quiescence[357], HSPC survival[358], influence leukocyte maturation and control megakaryocyte function[359]. Therefore alterations in both monocyte and monocyte-derived macrophage function in CMML could have a profound impact on patient phenotype.

As discussed above, monocytes in CMML have been implicated as major drivers of pro-inflammatory signalling through cytokine release[110]. In keeping with this

the proportion of classical monocytes, the most pro-inflammatory subtype, increases to >94% and atypical monocytes decrease to <1.13%[360], [361]. Despite recognised inter-patient heterogeneity, these findings have been externally validated by multiple studies and are now recognised as a diagnostic tool in the latest iterations of WHO and ICC classification criteria[362], [363]. CMML monocytes frequently express aberrant cell surface markers, with CD2 and CD56 being most specific for CMML, however the functional significance of this has not yet been elucidated[318], [364]. *Ex vivo* culture of CMML monocytes has found that they are more resistant to apoptosis than healthy donor monocytes via the activity of anti-apoptotic protein MCL1, which is propagated by CYTL1 autocrine signalling[95]. MCL1 has a dual function, in addition to its role in apoptosis it is responsible for regulating mitochondrial metabolism, promoting the use of oxidative phosphorylation[365]. Upregulated genes in CMML monocytes are enriched within pathways involved in oxidative phosphorylation and metabolomic analysis of respiratory capacity has found increased basal respiration, maximal respiration and spare respiratory capacity in keeping with enhanced oxidative phosphorylation[366]. Multiomic profiling of CMML monocytes has revealed that they are epigenetically dysregulated, with reduced accessibility of chromatin at genomic loci involved in mature myeloid cell function, indicating an immature phenotype. Furthermore, chromatin was less accessible at sites of NFκB binding, an important pro-inflammatory mediator discussed in the previous chapter, and analysis of paired gene expression data found no evidence of enhanced inflammatory cytokine expression as reported by *Franzini et al*[110], [366]. In fact, accessible chromatin was enriched for STAT6 binding, which is an important mediator of homeostatic (M2) macrophage polarisation[367]. Monocyte derived macrophages created by *ex vivo* culture of CMML monocytes exhibit increased expression on M2-associated genes and morphological characteristics compared to healthy donor controls suggesting they are primed for M2 polarisation [366]. Like monocytes, granulocytes in CMML have been found to be more immature, lacking CD16 expression, possessing fewer granules and having reduced phagocytic capacity to healthy controls[368]. Immature granulocytes, possessing similarities to

myeloid derived suppressor cells, have also been implicated by multiple studies as performing an important role in CMML pathogenesis through their release of resistin, alpha-defensins and CXCL8[111], [113], [366]. Therefore CMML is associated with plethora alterations to mature myeloid cell function, many of which have the capacity to alter HSPC function or explain the phenotypic spectra observed in CMML.

4.1.7 Summary and plan for investigation

CMML is associated with changes to the number and function of many different cell types. Monocytes are the pathognomonic expanded population and known to possess altered function. However, evidence is contradictory about whether they are pro-inflammatory or anti-inflammatory in nature. I will therefore use multiomic profiling to further characterise TET2m CMML monocytes both at bulk and single cell resolution in order to understand what influence these cells may exert on the bone marrow micro-environment. HSPCs reside within multiple niches in the bone marrow as they differentiate. I will map out the cellular composition of these niches using imaging mass cytometry and investigate whether the expansion of mature myeloid cell populations in TET2m CMML leads to altered cellular populations residing alongside HSPCs. Finally, I will use the single cell multiomic analysis of bone marrow mononuclear cells to predict which of these cell-cell interactions may have the greatest impact on HSPC dysfunction, which will be studied in further depth in the next chapter.

4.1.8 Aims

The specific aims for this section of my thesis are:

1. To characterise the spectrum of monocyte dysfunction in TET2 mutant CMML
2. To explore the changes in spatial demography of cell populations in TET2 mutant CMML bone marrow
3. To identify putative cell-cell interactions that may contribute to TET2 mutant CMML pathogenesis

4.2 Results

4.2.1 TET2 mutant CMML monocytes are transcriptionally and epigenetically distinct from healthy donor monocytes

TET2 dysfunction has been linked to dysregulation of inflammation[41], [162], [181] including pro-inflammatory changes to monocyte and macrophage function[38]. Current evidence of altered monocyte function in CMML is conflicting with some data supporting a pro-inflammatory change and other data a homeostatic change[110], [366]. To investigate this further in TET2m CMML, monocytes were sorted for RNA-seq and ATAC-seq analysis alongside the Lin$^-$ CD34$^+$ HSPCs discussed in the previous chapter (Figure 3.3A and B). Chromatin accessibility data was available for 15 TET2m CMML patients and 8 healthy donors. Gene expression data was available for 18 TET2m CMML patients and 10 healthy donors.

TET2m CMML monocytes have different chromatin accessibility to healthy donor monocytes but do not fully separate into HSPC defined phenotypic clusters

Monocytes from TET2m CMML patients had strikingly different chromatin accessibility within MACS2 called peaks than healthy donors. PCA analysis found that CMML samples separated from healthy donor samples along the PC1 axis (Figure 4.1A and C). Ward clustering identified five clusters, with two clusters made up of samples from healthy donors and three from TET2m CMML patients. There was no mixed cluster, as found in the analysis of HSPC samples. All monocyte samples had paired HSPC samples, sorted from the same bone marrow MNC vial, and therefore they could be annotated by their corresponding HSPC chromatin accessibility cluster. The majority of individuals, 6/7 erythroid-biased HSPC cluster 1 samples and 4/5 myeloid-biased HSPC cluster 2 samples, clustered together. Demonstrating that HSPC clusters remained relevant in clonally related monocytes and that HSPC phenotype indicates shared pathobiology (Figure 4.1B and C). However, these clusters were less distinct on PCA analysis of motif enrichment in open chromatin with samples becoming intermixed (Figure 4.1G and H). This

potentially indicates that global TF activity is more similar between individuals and that the differences between samples are found at individual loci.

Differential accessibility analysis was performed between HSPC defined clusters of monocyte samples to explore whether altered HSPC phenotype was associated with altered monocyte phenotype. Myeloid-biased TET2m CMML cluster 2 had 1417 differentially accessible peaks compared to healthy donors, but 66% of these peaks were also differentially accessible when comparing erythroid-biased TET2m CMML cluster 1 with healthy donor monocytes. This suggests a common TET2m CMML monocyte phenotype and therefore I conducted an analysis of the entire cohort of TET2m CMML monocytes and healthy donor monocytes. Differential accessibility analysis found accessibility was increased in 849 peaks and decreased in 4001 peaks (Figure 4.1D). GREAT gene ontology analysis of less accessible chromatin in TET2m CMML monocytes found enrichment for multiple pathways involved in myeloid leukocyte activation. These pathways are important for monocyte function and their reduced activity in TET2m CMML monocytes may indicate a deficit in their effector function or that they are less mature than their healthy donor counterparts.

The pattern of differential accessibility was different to that seen in HSPCs, where 5373 peaks were more accessible and 3756 less accessible in TET2m CMML compared to healthy donors. The position of these peaks were also different between DA peaks in monocytes and HSPCs. In monocytes 40% of more accessible peaks were in promoters and 27% in introns, and 53% of less accessible peaks were in introns and 9% in promoters. In HSPCs however, 50% of more accessible peaks were within introns and 13% promoters, and 37% of less accessible peaks were within introns and 13% promoters. The different pattern of changes in chromatin accessibility between HSPCs and monocytes may be a consequence of inherent differences in TF activity between these cell types. 60% of TFs exhibit altered binding affinity to methylated DNA and therefore cell type specific differences to TF activity, combined with global increase in methylation, particularly at enhancer and gene body loci, may explain these differences [146], [156]. To explore this hypothesis further I performed motif enrichment analysis of healthy donor samples

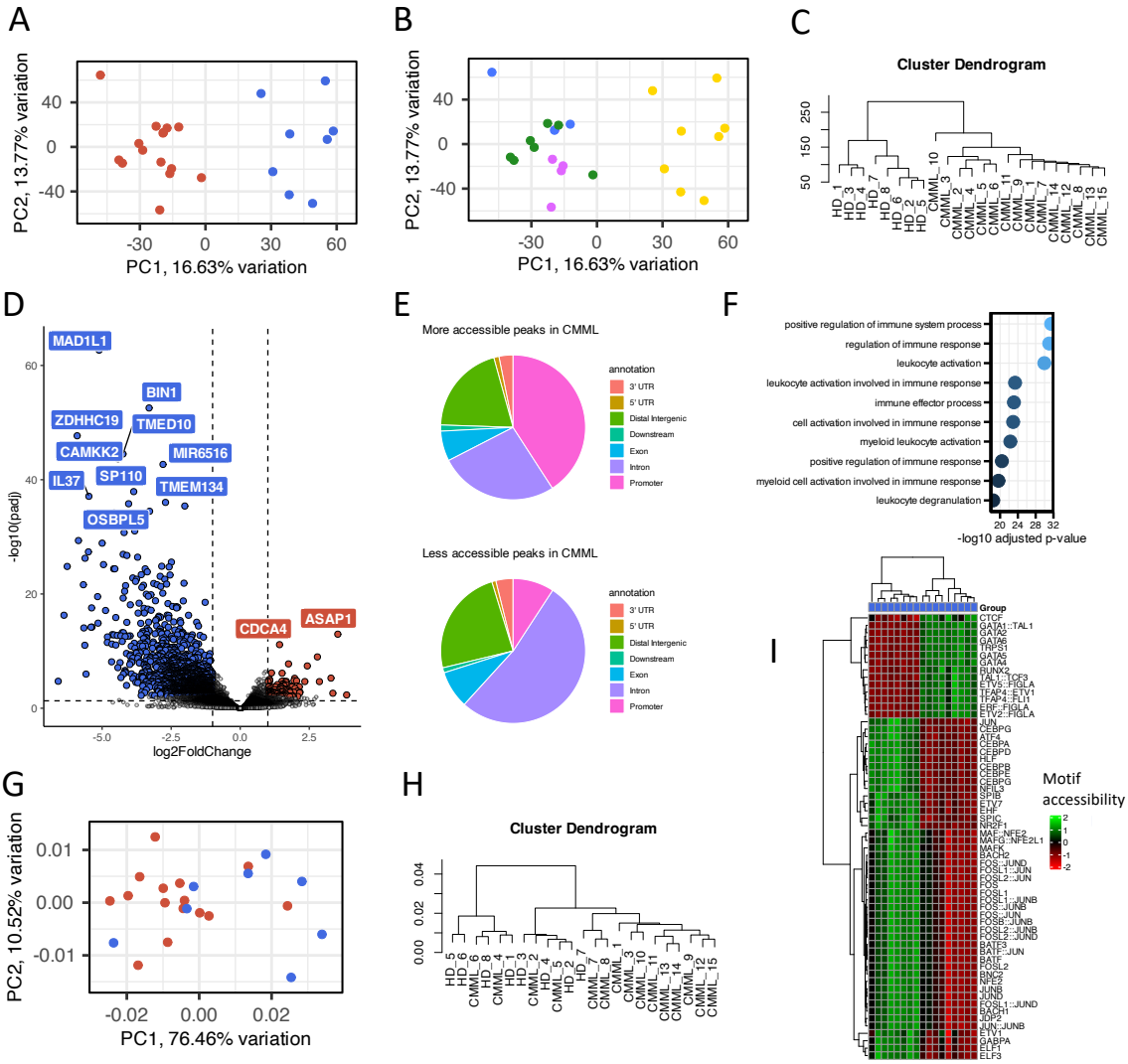


Figure 4.1: TET2m CMML monocytes have different chromatin accessibility to healthy donor monocytes but do not separate into HSPC defined phenotypic clusters - Principal component analysis of chromatin accessibility within reproducible high quality peaks coloured by disease group, red and blue for CMML and healthy donor samples respectively (A) and HSPC defined clusters, green, pink, blue and yellow for erythroid-biased cluster1, myeloid-biased cluster2, mixed cluster 3 and healthy donor cluster 4 respectively (B). (C) Ward clustering on principal components of chromatin accessibility. (D) Differential chromatin accessibility analysis between TET2m CMML monocytes and healthy donor monocytes. Labels for peaks were from nearest gene. (E) Annotation of the location of differentially accessible peaks between TET2m CMML monocytes and healthy donors. (F) GREAT gene ontology analysis of peaks with reduced accessibility in TET2m CMML monocytes. Principal component analysis of chromVAR motif activity scores, coloured by disease group (G) and ward clustering dendrogram (H). (I) ChromVAR motif activity score of most variable TFs between healthy donor HSPCs and monocytes.

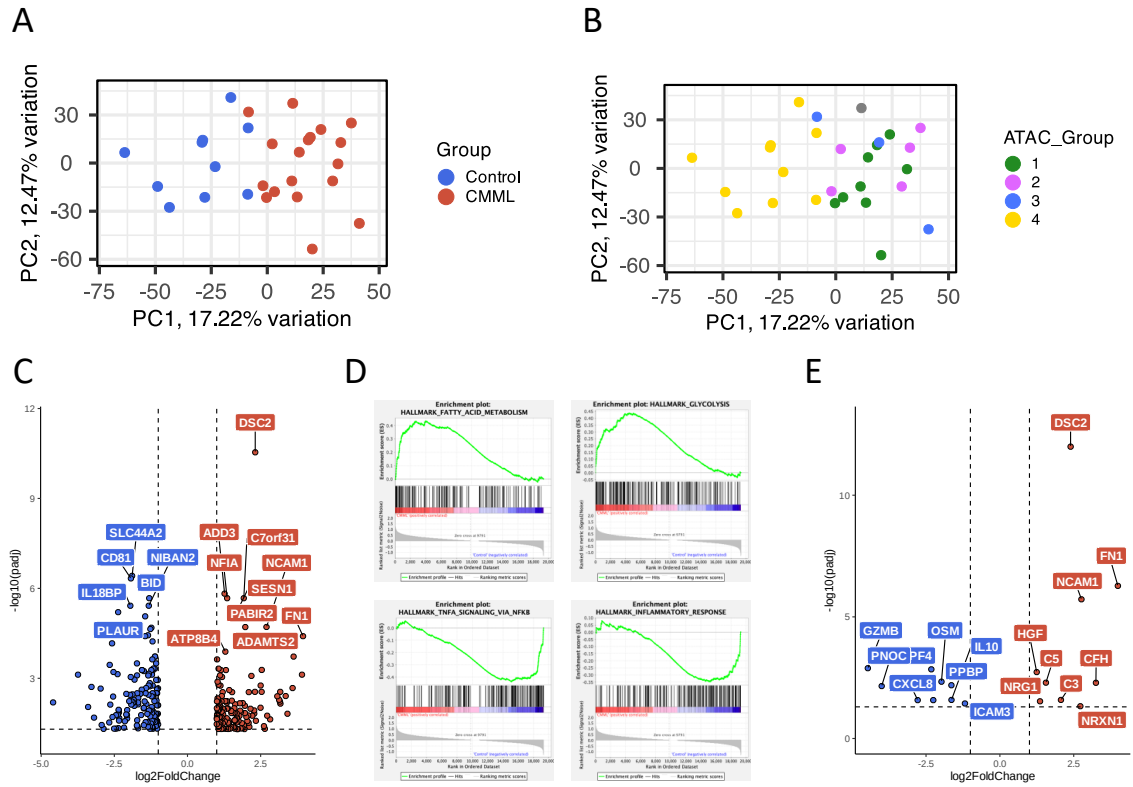


Figure 4.2: TET2m CMML monocytes have altered ligand expression to healthy donor monocytes - Principal component analysis of gene expression coloured by disease group, red and blue for CMML and healthy donor samples respectively (A) and HSPC defined clusters, green, pink, blue and yellow for erythroid-biased cluster1, myeloid-biased cluster2, mixed cluster 3 and healthy donor cluster 4 respectively (B). The TET2m CMML patient sample without paired ATAC-seq data is coloured in grey. (C) Differential gene expression between TET2m CMML monocytes and healthy donor monocytes. (D) Gene set enrichment analysis of MSigDB Hallmarks gene sets between TET2m CMML monocytes and healthy donor monocytes. (E) Differential ligand expression between TET2m CMML and healthy donor monocytes.

comparing HSPCs with monocyte samples to identify cell type specific TF activity. This found that FOS, JUN, ELF and BACH family motifs, which exhibit reduced binding to methylated DNA, were enriched in monocyte open chromatin, alongside C/EBP family motifs. Whereas GATA family motifs were enriched in HSPCs, of which the great majority of motifs do not include CpGs and are insensitive to methylation (Figure 4.1I). This may help to explain why more peaks were less accessible in TET2m CMML monocytes but more accessible in TET2m CMML HSPCs compared to healthy donors.

Gene expression analysis uncovers aberrant putative interactions between TET2m CMML monocytes and HSPCs

Principal component analysis of batch corrected gene expression data found that the greatest variance between samples was due to differences between TET2m CMML and healthy donors. Chromatin accessibility HSPC clusters did not separate TET2m CMML samples effectively, suggestive of a shared monocyte phenotype in TET2m CMML (Figure 4.2A and B). Differential gene expression analysis between TET2m CMML monocytes and healthy donor monocytes found 479 up-regulated and 630 down-regulated genes (Figure 4.2C). Gene set enrichment analysis found that TET2m CMML monocytes were enriched for fatty acid metabolism, glycolysis, E2F targets and G2M checkpoint pathways, indicating a metabolic switch to oxidative phosphorylation and increased proliferation. Healthy donor monocytes were enriched for inflammatory response and TNFA signalling via NFKB pathways. These pathways are required for mature monocyte function and their enrichment in healthy donor monocytes rather than TET2m CMML monocytes potentially indicates monocyte dysfunction or immaturity, just as was found in the GREAT gene ontology analysis of accessible chromatin reported above. Despite inflammatory signalling pathways being negatively enriched in TET2m CMML monocytes, some of the top differentially expressed genes were ligands, such as *DSC2*, *FN1* and *NCAM1*, all of which which have cognate receptors expressed on TET2m CMML HSPCs. Differential expression analysis using only a subset of published ligand genes [258] also found *HGF*, which has been implicated as a supportive factor in AML[369], and *NRG1* were upregulated and *IL10*, which suppresses blast proliferation in AML[370], *OSM* and *CXCL8* were down regulated (Figure 4.2E).

4.2.2 snRNA-seq facilitates identification of putative aberrant cell-cell interactions in TET2m CMML bone marrow

Subgroup analysis of mature myeloid cells alters cell type annotation and identifies two classical monocyte subsets

To further explore differences in mature myeloid populations in TET2m CMML I conducted a subcluster analysis of the single cell multiome dataset described in the previous chapter. I repeated integration of the mature myeloid cell subset using Harmony to remove variance from inter-individual differences. Clustering on this data identified that some cells had been mis-assigned during annotation of the complete dataset. I found a cluster of GM progenitors and another of CD4 T cells. I also found that monocytes could be separated into two groups based on *IL1B* expression, a marker of monocyte activation[371] (Figure 4.3A). Repeating integration, this time using Harmony to correct for batch alone, I found that mature myeloid cells, particularly GM progenitors and monocytes, separated cells into phenotypic groups. Pseudobulk differential gene expression analysis between erythroid-biased CMML and healthy donor monocytes (both *IL1B* high and low subsets were included) found 188 upregulated and 170 down regulated genes. A similar analysis between myeloid-biased CMML and healthy donor monocytes found 157 upregulated and 185 downregulated genes, with 109/358 (30%) of genes were shared between both analyses. Therefore, although differences between phenotypes can be detected on dimensionality reduction, the differences are subtle and many differences between CMML phenotypes and healthy donors are shared, similar to the bulk analysis. This can also be demonstrated through gene set enrichment analysis for MSigDb Hallmark gene sets of gene lists ranked by Wilcoxon AUC between erythroid or myeloid-biased CMML and healthy donor monocytes (Figure 4.3 E and F). Top positively and negatively enriched gene sets are shared between analyses, with CMML monocytes being enriched for interferon response gene sets and negatively enriched for MYC target and E2F target gene sets, suggesting they are less proliferative but more pro-inflammatory than healthy donor monocytes.

Surprisingly, some of these gene sets display opposite enrichment to the analysis of bulk gene expression in monocytes, as shown in Figure 4.3D. However, this difference may potentially be explained by an increase in the number of GM progenitors in TET2m CMML, although the difference in the proportions of GM progenitors in total BM-MNC did not reach statistical significance (p=0.2) due to inter-individual variance. GM progenitors also express CD14 and therefore will be found in the bulk samples analysed. Subgroup analysis comparing GM progenitors with monocytes in all individuals found enrichment for E2F targets, G2M checkpoints and Oxidative Phosphorylation gene sets with negative enrichment for Inflammatory response and TNFA signalling via NFKB gene sets. Therefore changes in enrichment for these gene sets in the bulk analysis is likely to be due to differences in cell type composition of each sample rather than altered function of specific cell types.

Somatic TET2 mutations originate within haematopoietic stem cells in TET2m CMML. This results in reduced ability to remove CpG methylation at TET2 specific target loci and this methylation will be inherited by all daughter cells including terminally differentiated monocytes. It is therefore possible that methylation sensitive TFs could have persistent altered function throughout differentiation. To test whether I could identify any shared changes to TF gene expression and motif enrichment between HSC and monocytes I conducted a paired analysis of gene expression and chromatin accessibility of monocytes in a similar way to the analysis I performed on HSCs in the previous chapter (Figure 3.14C and Figure 4.3G). This did not find any overlap in TFs with greatest average Wilcoxon AUC, which is perhaps not unexpected because different TFs are required to coordinate gene regulation at different time points throughout differentiation. However, NFKB1, NFKB2 and REL activity was downregulated in CMML monocytes compared to healthy donor monocytes. This is interesting, as each of these genes code for members of the NF-κB complex, a master regulator of inflammation[372]. The role of NF-κB in coordinating inflammation is complex and multifaceted, with different functions being described in different immune populations. However NFKB1 loss of function variants are the most prevalent cause of common variable immunodeficiency in Europeans,

potentially implicating this difference as a source of acquired immunodeficiency seen in CMML patients[373]. Furthermore, *NFKB1* knockout in THP-1 derived macrophages causes increased IL-6, IL-1β and TNF production, but impaired T cell co-factor expression[374]. Reduced NF-κB function in THP-1 derived macrophages therefore resulted in the paradoxical situation of increased pro-inflammatory cytokine production coupled with reduced coordination of the adaptive immune response, a situation similar to that seen in CMML associated with systemic inflammatory and autoimmune disease (SIADs)[4].

Pseudotime analysis of *CEBPB* indicated increased motif enrichment and RNA expression in CMML monocytes (Figure 3.14D and E). To explore the gene regulatory changes responsible for this I visualised coverage tracks of the *CEBPB* loci and found that CMML monocytes had increased chromatin accessibility at both the *CEBPB* promoter and a monocyte specific distal enhancer site[173] (Figure 4.3J). C/EBP-β therefore represents a TF with enhanced activity throughout monocyte differentiation in TET2m CMML.

4.2.3 Classical monocyte populations are expanded 10 fold in TET2m CMML bone marrow mononuclear cells and exhibit an aberrant immunophenotype

CMML is defined by an increase in classical monocyte production[360]. To explore whether additional mature myeloid populations are expanded in TET2m CMML I characterised the changes in abundance and immunophenotype of mature mononuclear phagocytic cell populations within the bone marrow of TET2m CMML patients. Classical monocytes were 10 fold more abundant in TET2m CMML bone marrow mononuclear cells than in healthy donors (Figure 4.4A). Classical monocytes also expressed more CD64 but reduced CD11c and CD45RA than healthy donor monocytes on average. However, there was variability between individuals. Classical monocytes of some CMML patients also aberrantly expressed CD123, but this was not significantly different on average between groups (Figure 4.4B). These findings are in keeping with previous studies of aberrant marker expression in CMML [318], [360], [375]. Intermediate monocytes were also increased in TET2m CMML,

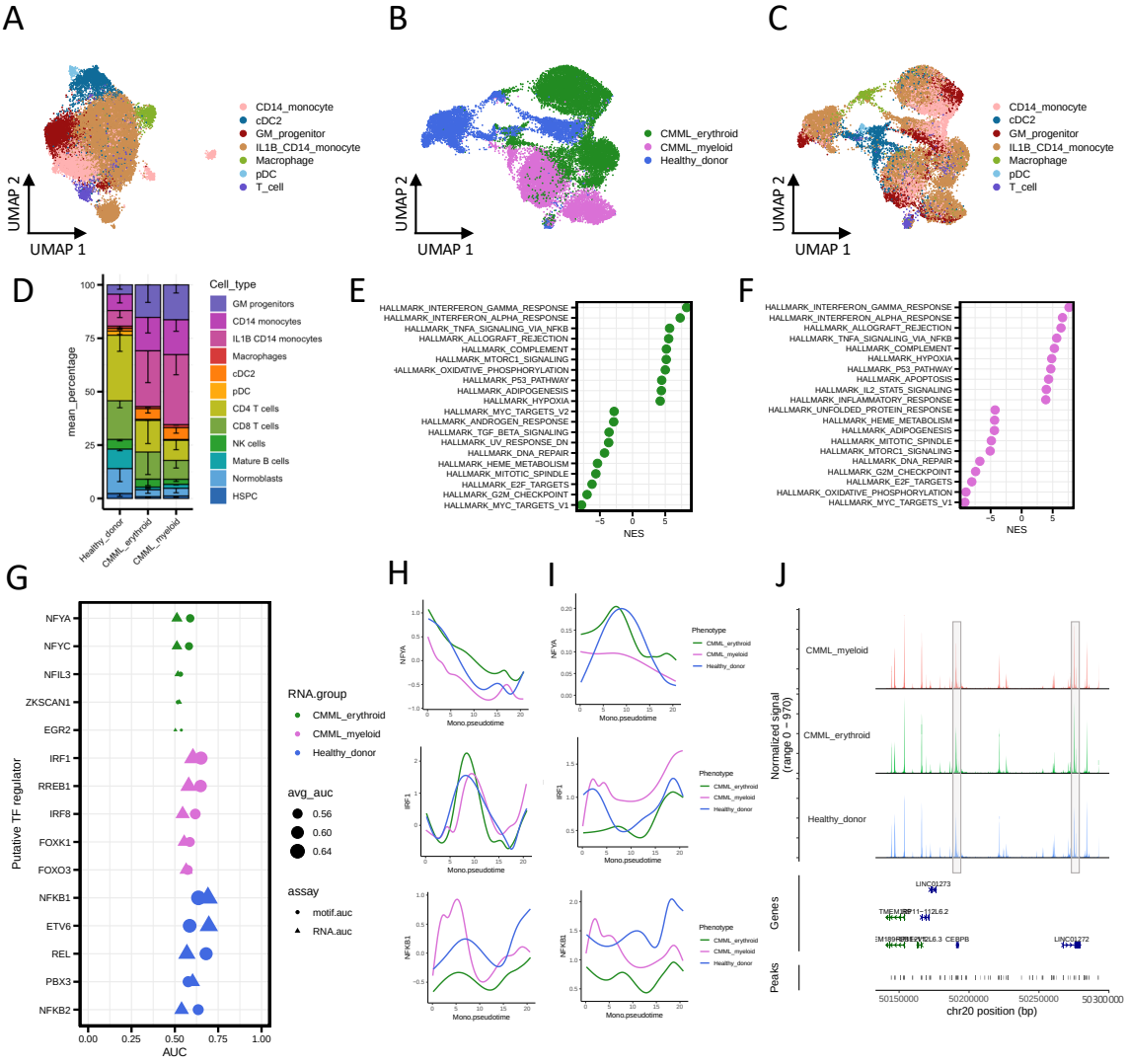


Figure 4.3: Subgroup analysis of mature myeloid cell populations improves accuracy of cell type annotation and identifies differences as well as similarities between CMML phenotypes - (A) UMAP projection showing revised cell annotation of mature myeloid cell subgroup after repeat WNN integration of RNA and ATAC data that was adjusted for inter-individual variance using harmony. UMAP projection for WNN integrated RNA and ATAC data that was adjusted for batch variance using harmony showing retained differences between phenotypic groups (B) that is most apparent in monocyte populations (C). (D) Average percentage of mature myeloid cell types for each individual displayed by phenotypic group. Data is from complete dataset rather than subgroup analysis. Error bars show standard error of means. GSEA analysis using MSigDb Hallmarks gene sets of differentially expressed genes between erythroid-biased CMML and healthy donor monocytes (E) and myeloid-biased CMML and healthy donor monocytes (F). (G) Top putative transcription factor regulators identified by Wilcoxon AUC test of each phenotypic group against the other two phenotypic groups for gene expression (triangle) and motif enrichment (circle) analysis. Top putative TF regulators plotted by motif enrichment over monocyte differentiation pseudotime (H) and normalised gene expression (I). (J) Chromatin accessibility around CEBPB locus with promoter and distal regulatory region highlighted.

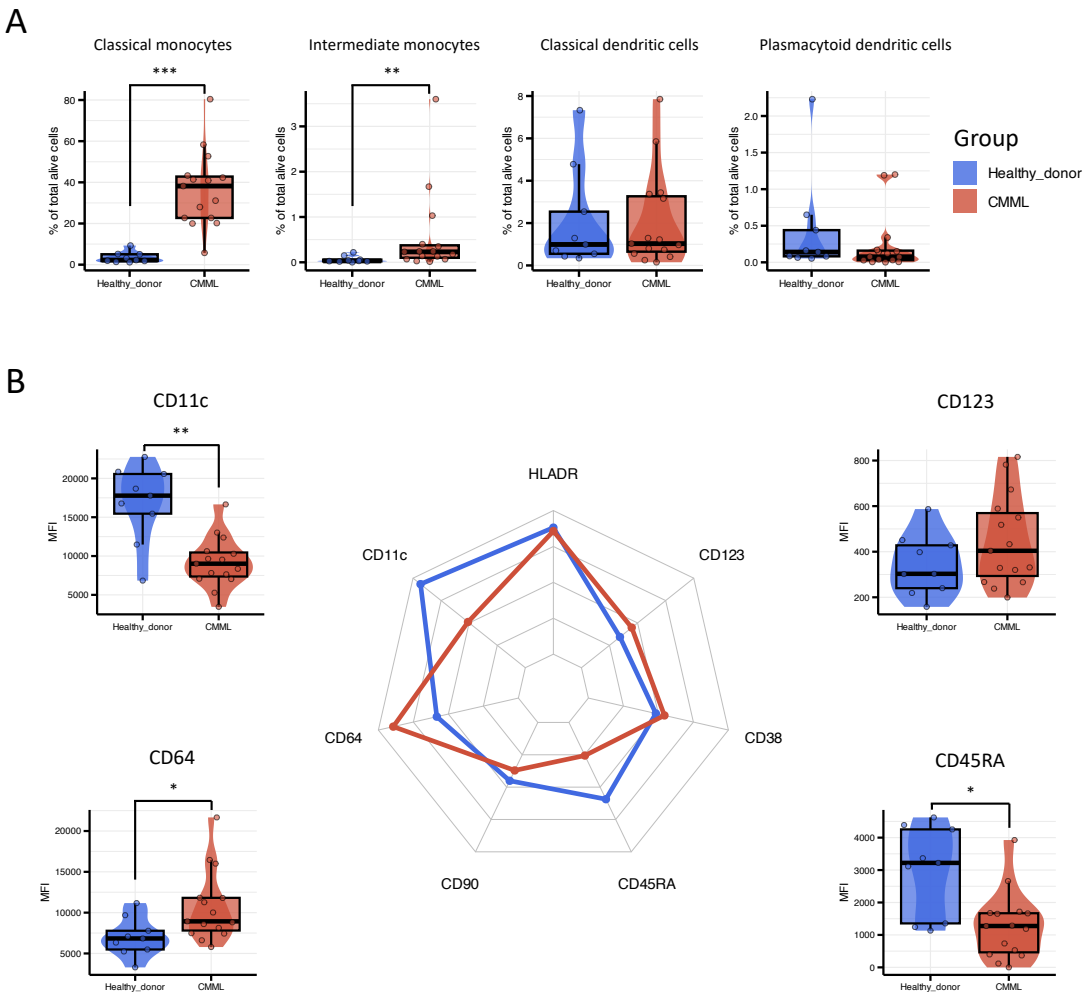


Figure 4.4: Changes in abundance and immunophenotype of mature myeloid populations in TET2m CMML - (A) Percentage of alive cells for classical monocytes (HLADR$^+$ CD14$^+$ CD16$^-$ative), intermediate monocytes (HLADR$^+$ CD14$^+$ CD16$^+$), classical dendritic cells (HLADR$^+$ CD14$^-$ CD16$^-$ CD11c$^+$) and plasmacytoid dendritic cells (HLADR$^+$ CD14$^-$ CD16$^-$ CD123$^+$). (B) Classical monocyte aberrant immunophenotype, median fluorescent index (MFI) displayed as a spider plot and for specific markers of interest. CMML samples in red and healthy donor samples in blue. Groups were compared by Wilcoxon test and all significant ($p<0.05$) differences are displayed with brackets

although this difference was driven by a small number of individuals with higher CD16 expression. This phenotype has been previously identified in CMML patients with an associated inflammatory state such as caused by a concurrent autoimmune disease[376]. Neither plasmacytoid dendritic cells nor classical dendritic cells were present at a different abundance in TET2m CMML.

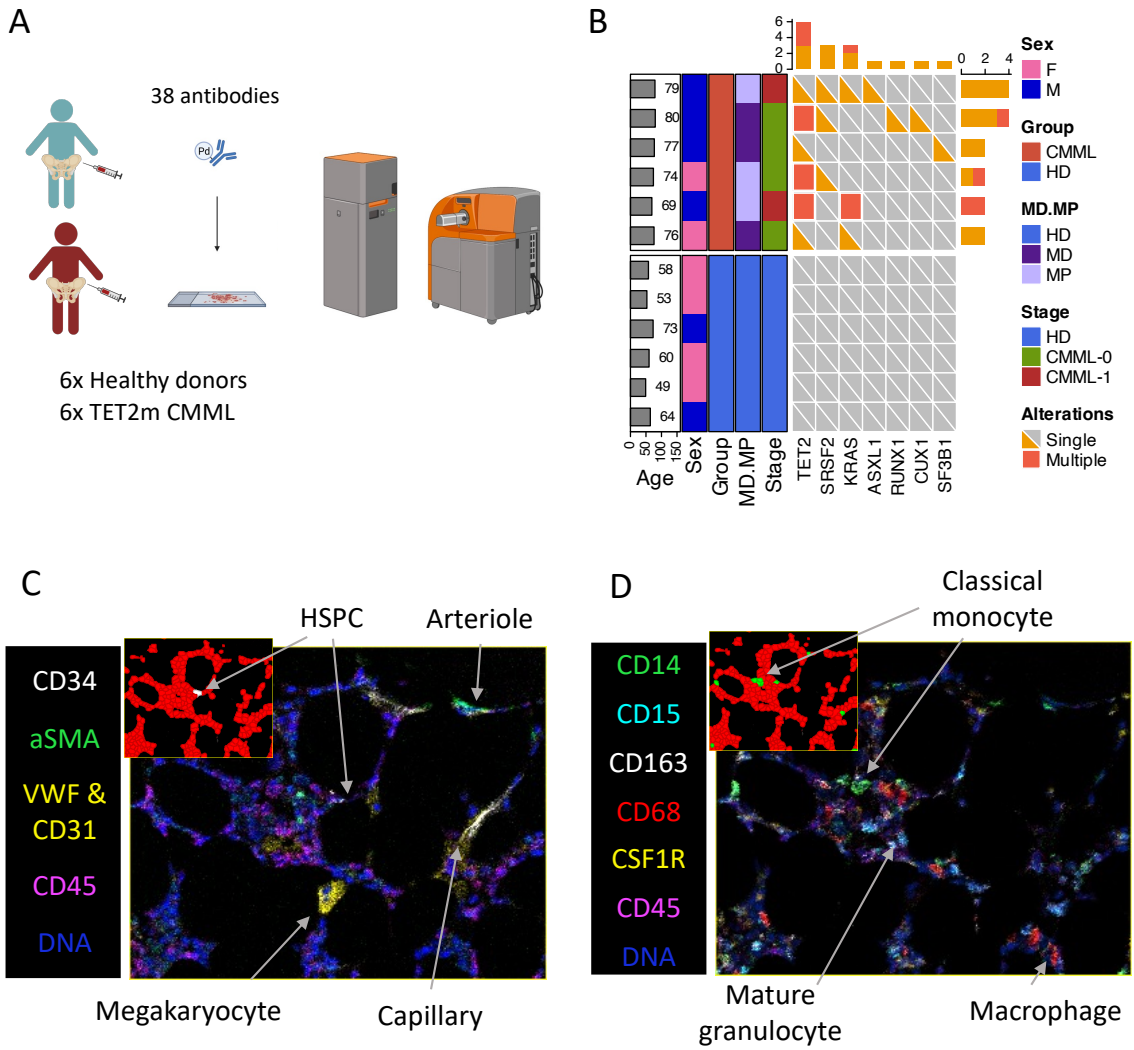


Figure 4.5: Overview of imaging mass cytometry experiment - (A) Summary of imaging mass cytometry experiment. (B) Description of demographic details of individuals included in the imaging mass cytometry experiment. Representative images of a healthy donor trephine sample highlighting how HSPCs can be distinguished from megakaryocytes and blood vessels (C) and classical monocytes from granulocytes and macrophages (D). Inserts display cell segmentation masks of the same area with HSPCs (C) and classical monocytes (D) highlighted in yellow and green respectively and all other cell types in red.

4.2.4 TET2m CMML is associated with an increase in both classical monocyte and macrophage populations in the bone marrow

To explore changes to cell population abundance and topography in TET2m CMML I performed imaging mass cytometry on bone marrow trephine samples from 6 healthy donors and 6 TET2m CMML patients (Figure 4.5 A and B). 24 regions of

interest were identified using haematoxylin and eosin stained images of consecutive sections. Areas of high quality tissue that represented the tissue as a whole were selected. Images were acquired on the Standard BioTools Hyperion imaging mass cytometry (IMC) instrument. Analysis was performed using the steinbock pipeline for image pre-processing and hot pixel removal[260]. Pixel classification and image segmentation was performed in ilastik[261]. Images and segmentation masks were exported to QuPath for quality control and multiplex image visualisation[262]. Images were of good quality with excellent staining in most channels. However, staining in CD56, CD123 and IL-1β channels were weak and therefore only the cells with strongest expression could be detected. Thresholds were applied to each channel to identify positive and negative marker expression either by setting a threshold manually, or by building a random forest classifier. 198560 cells were identified by cell segmentation, made up of 88673 healthy donor cells and 109887 cells from TET2m CMML patients. 47% of cells could be classified to a cell type using Boolean gating for canonical cell type markers (see Methods section for details of gating strategy). 21 cell type markers were included in the panel and therefore it was not possible to confidently identify all the cell populations present within the bone marrow. Stringent thresholds were used to prevent false positive cell type assignment, which reduced the total number of cells assigned a cell type. Cell segmentation for megakaryocytes was poor, with clear evidence of oversegmentation and therefore, although it is possible to identify megakaryocytes in images, they have not been included in the broader analysis. Likewise, adipocytes were excluded because the large central fat spaces of these cells were not identified as intracellular by cell segmentation algorithms. HSPCs are challenging to identify in the bone marrow because CD34, the canonical phenotypic marker is also expressed on endothelial cells of blood vessels. CD45 can be used to distinguish these cell types as it is ubiquitously expressed specifically in haematopoietic cells, however bleeding of pixels between adjacent cells can make it difficult to confidently assign cells to one population rather than the other. To combat this I included additional markers: CD31, which is expressed on venous endothelial cells, and alpha-smooth muscle actin (aSMA), which

identifies smooth muscle cells surrounding arterioles. By incorporating all of these markers into a random forest classifier I was able to successfully identify HSPCs from endothelial cells. Correct cell type annotation of all cell types was confirmed by visualisation of images and masks from each region of interest (Figure 4.5c and D). Furthermore, cells that were not annotated were uniformly distributed throughout tissue sections and between regions of interest, providing reassurance that the incomplete cell type annotation was not introducing excess bias to the analysis.

Cell density was increased in TET2m CMML patients compared to healthy donors (8155 vs 5559 cells per mm^2, p = 0.003, Figure 4.6C). The proportion of cell types within the bone marrow of healthy donors and TET2m CMML was calculated using the average percentage of each cell type by individual (Figure 4.6C to F). This found a 3 fold increase in classical monocyte percentage (7.6% vs 22.9%, p = 0.02) and a trend towards an increase in macrophage percentage in TET2m CMML (7.2% vs 12.6%, p=0.06). Interestingly there was a 4 fold decrease in the percentage of mature $CD15^+$ granulocytes in TET2m CMML (16.3% vs 3.6%, p=0.009) suggesting there is a failure of terminal differentiation of neutrophils in TET2m CMML. HSPC percentages were no different between groups, despite two CMML patients having been classified as CMML-1 (5-10% blasts) on bone marrow aspirate analysis. However, poor correlation between blast counts on trephine samples and aspirate smears can be seen when blast percentages are <10%[377]. Furthermore, promonocytes, which resemble blasts on morphological assessment, are included in the blast count when staging CMML. However, promonocytes do not express CD34 and therefore were not possible to identify by IMC. Comparison of cell density by calculating average cells per mm^2 found similar patterns of changes in cell density to cell percentage (Figure 4.6H), confirming the validity of cell type annotation. During cell type annotation I identified a CSF1R positive population that appeared to be macrophage-like by morphology, but did not consistently express either CD68, CD163 or HLADR (Figure 4.6A). I therefore annotated this population as CSF1R macrophages, but they may well belong to an alternative CSF1R expressing haematopoietic cell type, such as a monocyte/macrophage

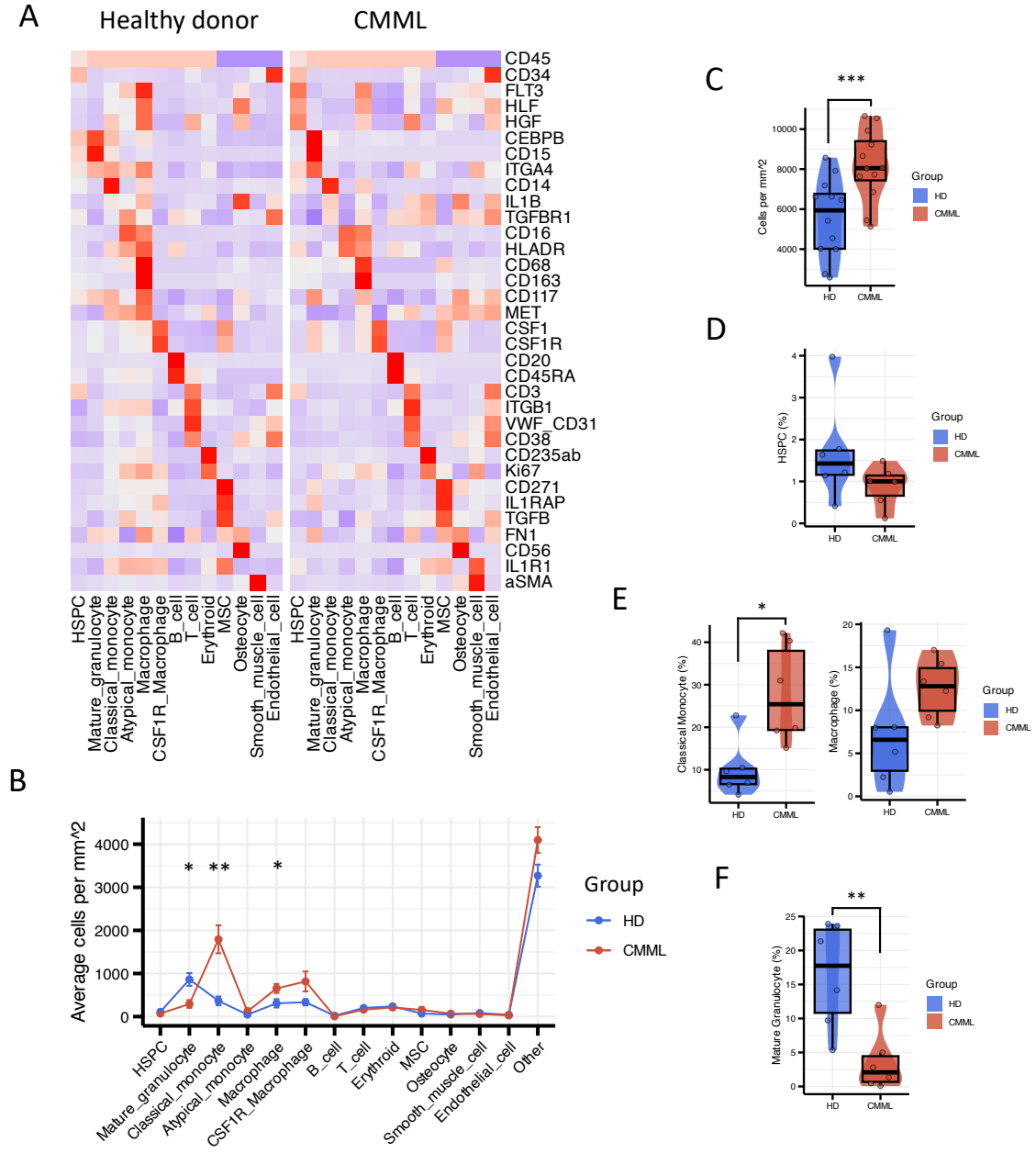


Figure 4.6: Imaging mass cytometry confirms that TET2m CMML is associated with an expansion of classical monocyte and macrophage populations - (A) Scaled cell marker expression by cell type. Red indicates increased expression and blue indicates decreased expression. (B) Average cell number per mm^2 of all cell types. Blue line displays healthy donor and red line TET2m CMML cells. Error bars demonstrate standard error of means. (C) Average cell number per mm^2 by individual. Average percentage of total cell number by individuals for HSPCs (D), classical monocytes (E), macrophages (F) and mature granulocytes (G). Differences between groups were tested by Wilcoxon test. All significant ($p<0.05$) differences are displayed using brackets.

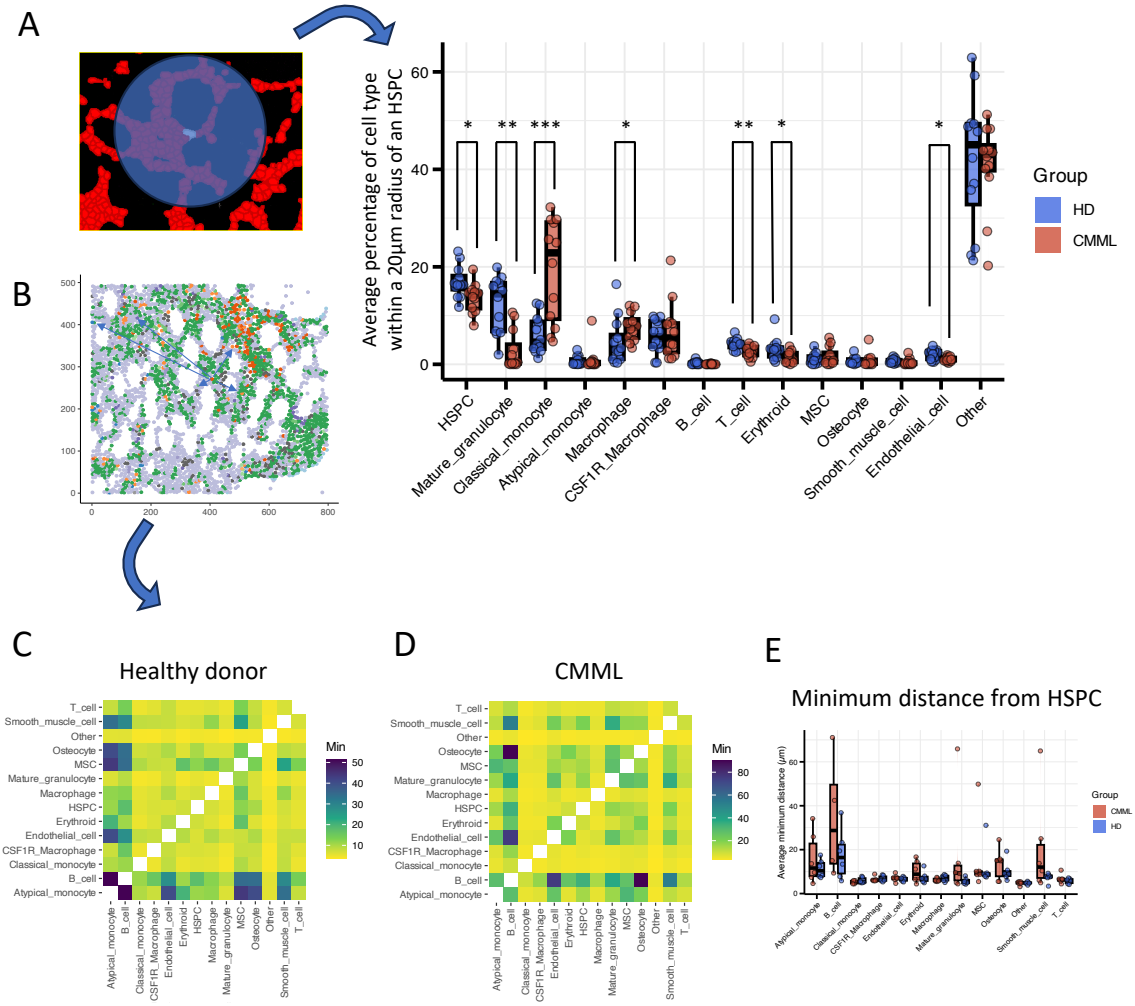


Figure 4.7: Spatial analysis uncovers alterations to the cells within the HSPC niche in TET2m CMML - (A) Average percentage of cell types within a 20μm diameter of each HSPC calculated using SPIAT, plotted by individual within disease group[263]. (B) Schematic representation of calculation performed to identify minimum distance between one cell type and another for each region of interest. Mean minimum distances in μm between individuals were calculated and plotted in heatmaps for healthy donors (C) and CMML (D). (E) Average minimum distances between HSPCs and other cell types plotted by disease group. Differences between groups were tested by Wilcoxon test. All significant ($p<0.05$) differences are displayed using brackets.

precursor. This cell population was expanded within some TET2m CMML patients, although overall there was no statistical difference in density to healthy donors (797 vs 294 average cells/mm^2, p=0.31 by Wilcoxon test).

4.2.5 HSPCs reside in close proximity to classical monocytes and macrophages in TET2m CMML bone marrow

I performed spatial analysis using SPIAT to compare cell type composition of HSPC niches between TET2m CMML and healthy donors[263]. I identified cell types that were in closest proximity to HSPCs by calculating the minimum distance between HSPCs and all other cell types for each individual and averaged them by group (Figure 4.7B to E). This found that HSPCs can reside close to most cell types only atypical monocytes, B cells, osteocytes and smooth muscle cells had an average minimum distance of over 10μm from HSPCs. There were no significant differences between disease groups in minimum proximity to HSPCs for any cell type. However, this analysis does not account for changes in the abundance of cell populations within the HSPC niche. To compare this I performed a quadrat analysis centred on each HSPC and calculated the average percentage of each cell type residing within a 20μm diameter for each region of interest and compared these averages between healthy donor samples and TET2m CMML samples (Figure 4.7A). This found significantly increased classical monocytes and macrophages within the TET2m CMML HSPC niche, and significantly fewer HSPCs, mature granulocytes, T cells, erythroid cells and endothelial cells. These data confirm our hypothesis that the HSPC niche is disrupted within TET2m CMML and therefore interactions between classical monocytes and macrophages are likely to have more of an impact of HSPC function. Therefore, by studying these interactions it may be possible to identify novel pathways to target as treatment for CMML.

4.2.6 Multinichenet analysis identifies putative interactions that are predicted to promote HSPC dysfunction in TET2m CMML

CMML is associated with a dramatic increase in bone marrow cellularity, particularly because of expansion of myelomonocytic precursors and mature monocytes. Haematopoietic lineages, which are usually topologically distinct, become jumbled and in most patients there is evidence of overt dysplasia[378]. These profound

changes to the bone marrow micro-environment undoubtedly alter HSPC niches, leading to aberrant cell-cell interactions with cell types that are usually not present within their niche. To explore these altered HSPC interactions I applied multinichenet to the single cell multiomic dataset. This tool uses a multifaceted approach to identify putative cell-cell interactions. Not only does it identify interactions where sender and receiver cells express cognate ligand-receptor pairs, but it also allows comparison of different phenotypic groups to try to identify interactions that are predicted to impact gene expression in differentially expressed genes between these groups. The top prioritised putative cell-cell interactions are displayed by circus plots in Figure 4.8A. In particular I focused on interactions where the receiver cells were either HSC/MPPs or CMP/GMPs because these interactions were most likely to help propagate the stereotypical phenotype of CMML. Interactions between HSPCs were included, but in practice it is not common to see large aggregates of HSPCs in chronic phase CMML, and therefore these interactions are less likely to be significant for pathogenesis. Therefore the most important interactions are likely to be between HSPCs and mature myeloid populations, lymphocytes or erythroid cells. After reviewing the pattern of interactions, several ligand-receptor interactions were recurrently identified, or involved in pathways known to be important for regulating haematopoiesis or myeloid malignancy. I used these interactions as a basis for creating a short list of potential target interactions to further validate as being drivers of TET2m CMML, which are shown in Figure 4.9.

TGF-β signalling is known to be important in regulating HSC quiescence and negatively regulating myelopoiesis[379], [380]. Multinichenet predicts that TGF-β ligand activity is reduced in erythroid-biased CMML and therefore downregulation of this signalling pathway may promote the development of CMML.

Differential gene expression analysis found *IL1RAP* to be increased in CMML HSC/MPPs and *IL1B* ligand activity was increased in myeloid-biased CMML in the multinichenet analysis. IL1RAP has previously been reported as having increased expression in CMML, which further increased on blast phase transformation[15]. It has also been implicated in driving AML via dimerisation with c-KIT and

FLT3 receptors[381]. IL-1β is a pro-inflammatory cytokine, which is involved in supporting TET2m clonal haematopoiesis during aging[184], [382] and is emerging as a possible therapeutic target in myelodysplastic syndromes[383]. Therefore, I chose to investigate IL1RAP signalling further via potential partnership with either IL1R1, c-KIT or FLT3.

Macrophage colony stimulating factor (M-CSF or CSF1), released by normoblasts, was found to potentially interact with CSF1R, CSF2RA and CSF3R on CMP/GMPs by multinichenet in erythroid-biased CMML. *CSF1R* was also expressed, but at a very low level, in HSCs and MPPs. However, expression was below the threshold used for multinichenet and therefore, although M-CSF/CSF1R interactions are likely to be possible with this subpopulation, they were not highlighted as an interaction of interest by the algorithm. M-CSF is an important regulator of myelomonocytic differentiation of HSPCs[269] and is required for survival of monocyte and macrophage populations[384]. This interaction is particularly interesting, as it would help link the accumulation of MEPs and erythroid progenitors seen in erythroid-biased CMML with increased myelomonocytic cell production. Furthermore, CSF1R is a downstream target of C/EBP-β[385], therefore M-CSF release by erythroid populations may be one of multiple potential signals that helps facilitate differentiation of TET2m CMML HSPCs primed by C/EBP-β for myelomonocytic differentiation.

Neural cell adhesion molecule 1 (NCAM1, CD56) is aberrantly expressed by classical monocytes in 47-80% of CMML patients [360], [386]. Review of *NCAM1* expression in the multiome dataset found that some CMP, GMP, GM progenitors and cDC2 cells also aberrantly expressed *NCAM1* in CMML. NCAM1 expression on the stromal cell line FMS/PA6-P and mesenchymal stem cells (MSC) has been found to support HSPC proliferation and long term colony forming capacity[96], [387]. The receptor responsible for this was not elucidated, but reports from other cell types has found NCAM1 can act as a receptor after binding to itself or other ligands such as TAG-1 and the GDNF family[388]–[390]. NCAM1 can also act as a ligand to

fibroblast growth factor receptor-1 (FGFR1), leading to AKT/ERC/SRC signalling activation in HeLa cells, as well as performing a vital role in cell adhesion[391], [392].

Galectin-9 (LGALS9) was identified as a prioritised ligand by multinichenet, particularly in myeloid-biased CMML. Galectin-9 has been implicated in altering tumour micro-environment in solid organ malignancies to enhance activity of immunosuppressive T helper cells and myeloid suppressor cells[393], which have recently also been identified as important players in the aberrant bone marrow niche in CMML[111]. However the direct interaction between galectin-9 and HSPCs has not been well established, and the receptors identified by multinichenet were not part of the canonical pathway, where galectin-9 interacts with TIM-3. Therefore, I opted not to investigate this interaction further at present.

Fibronectin (FN1) and hepatocyte growth factor (HGF) were also included as ligands of interest even though they were not identified by multinichenet as a prioritised interaction. *FN1* was found to be one of the most significantly overexpressed ligands in the bulk RNA-seq analysis of TET2m CMML monocytes. Fibronectin has been found to improve ex vivo haematopoietic stem cell expansion and help retain HSC regenerative capacity[332], [394]. Its cognate receptor VLA-4 (integrin $\alpha 4$ $\beta 1$) has also been implicated in supporting AML cell survival and therefore targeting this axis may be of benefit in CMML[395]. HGF, which interacts with the receptor c-MET, is an important growth factor in supporting HSPC colony formation in *in vitro* models, functioning in a synergistic manner with IL-3 and GM-CSF[396], [397]. HGF is predominantly produced by mesenchymal stem cells in the bone marrow but analysis of the multiome data found widespread expression in GMPs, GM progenitors and monocytes, particularly in erythroid-biased CMML[398]. HGF has also been identified as a potential therapeutic target in AML, malignant blasts in 40% of AML patients produce HGF, leading to autocrine HGF-MET signalling that can support blast survival and proliferation[369], [399].

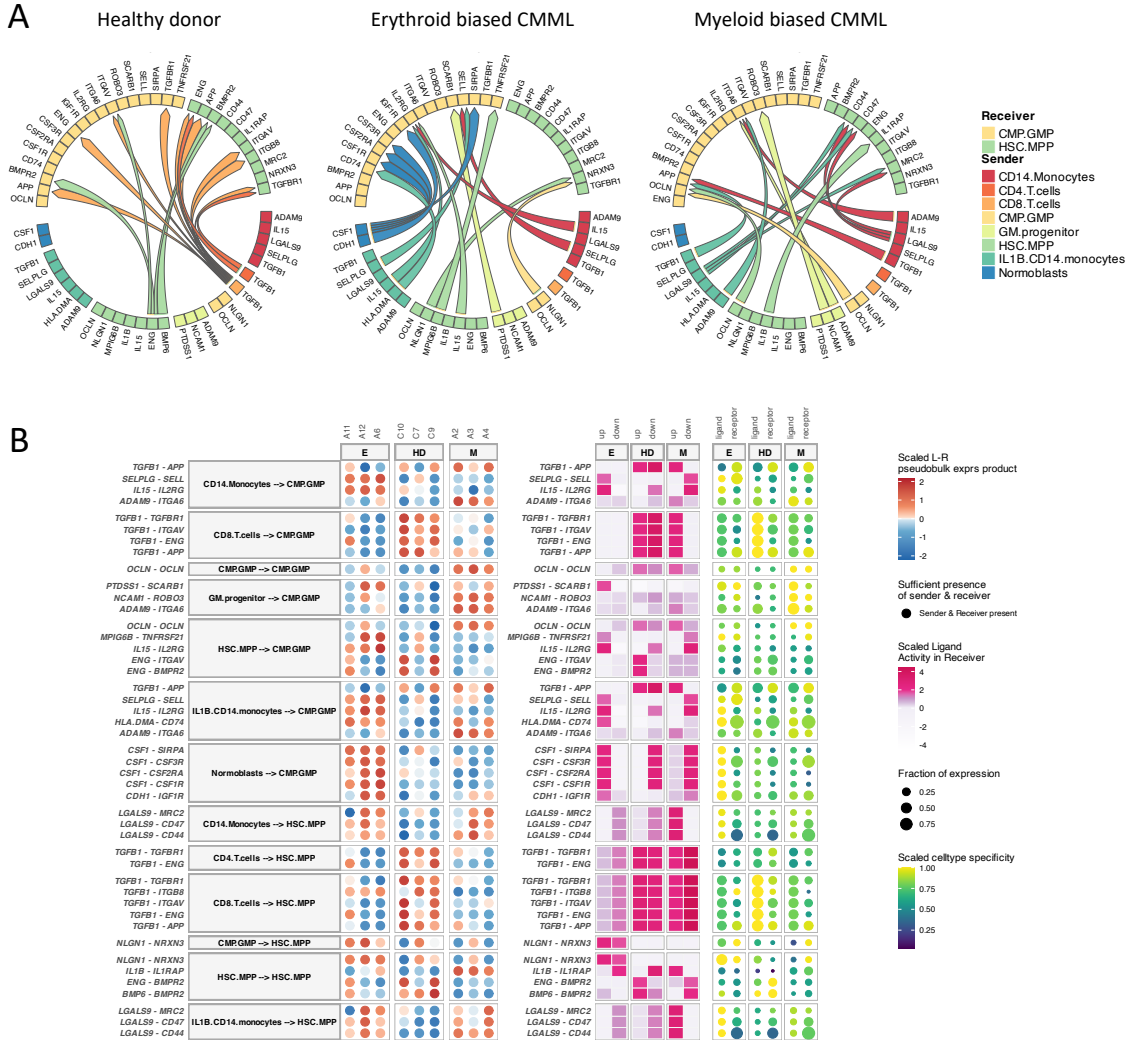


Figure 4.8: Multinichenet analysis identifies putative interactions predicted to support HSPC dysfunction in TET2m CMML - (A) Circus plots of prioritised ligand-receptor interactions between major identified haematopoietic sender cells and CMP/GMP or HSC/MPP receiver cells. (B) Heatmap of top prioritised interactions displaying scaled product ligand and receptor expression by individual (left), scaled predicted ligand activity and cell type specificity (right).

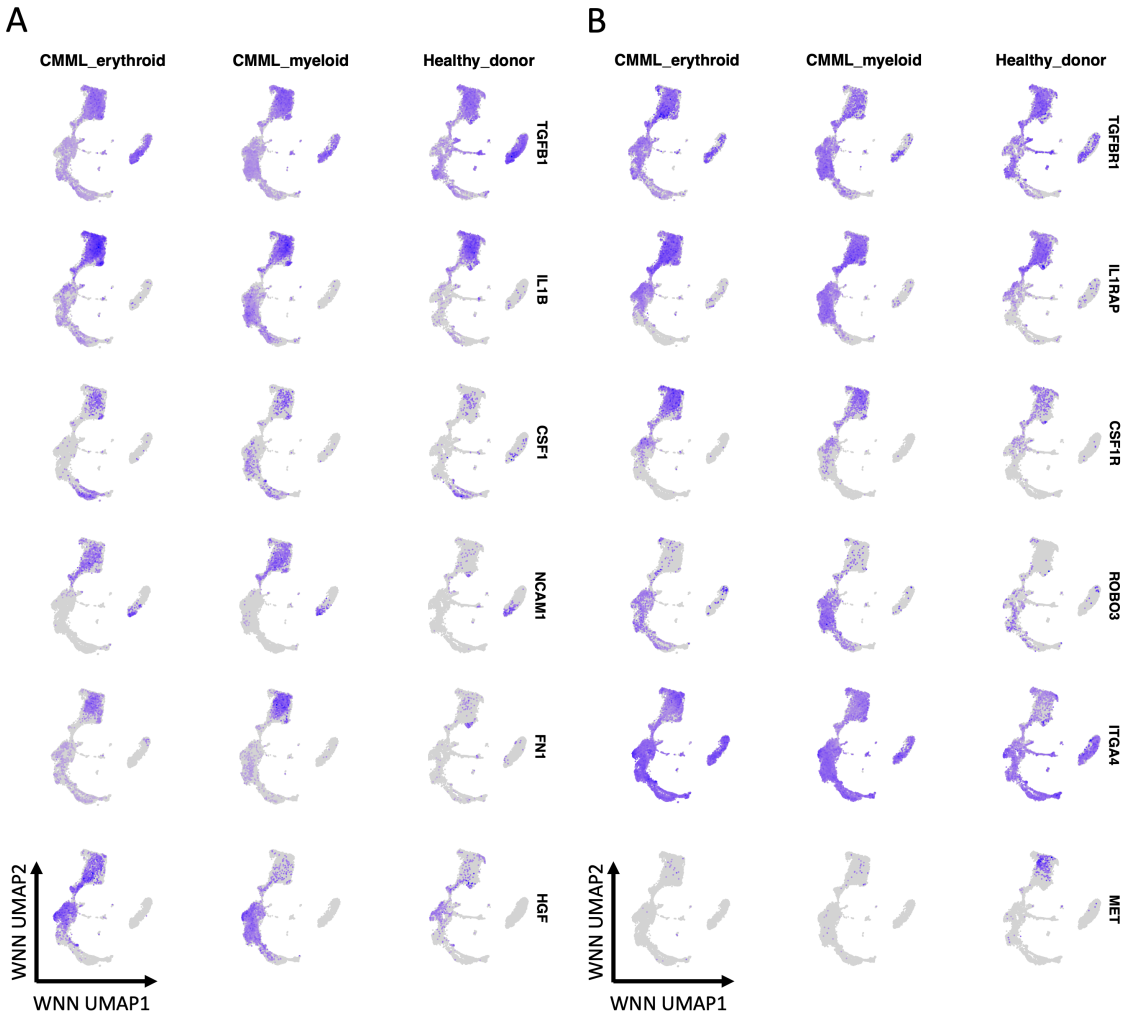


Figure 4.9: RNA expression of candidate ligands and receptors split by phenotypic group - Candidate ligand (A) and cognate receptor (B) expression in multiome dataset. Gene name is on right, expression displayed in cells from erythroid-biased CMML (CMML_erythroid), myeloid-biased CMML (CMML_myeloid) and healthy donors (Healthy_donor). Cell-type distribution on UMAP is shown in Figure 3.10C.

HSPCs express multiple receptors of interest but there was no difference in expression between healthy donors and TET2m CMML

Using the bone marrow trephine IMC data, HSPCs were subset and the percentage of HSPCs expressing specific receptors above threshold were calculated for each individual. There were no significant differences in IL1RAP, TGFBR1, CSF1R, NCAM1 and MET receptor percentage expression between TET2m CMML patients and healthy donors. Receptor expression of IL1RAP partners was generally low (<5%). Surprisingly average IL1RAP expression was low in all individuals and

there was a trend towards a reduction in CMML (4.9% vs 1.5%, p=0.06). This was the opposite to the expected trend from gene expression data. TGFBR1 also exhibited an opposite trend to gene expression, where expression trended towards being increased in CMML compared to healthy donors (19.2% vs 4.33%, p=0.06). Average CSF1R expression was 16.6% and 16% in healthy donors and CMML patients respectively. NCAM1 expression was low in both groups, although the antibody signal was weak and therefore only very high expression could be detected. MET signal however was strong, but average expression was 4.6% and 5.9% for healthy donors and CMML patients respectively. Integrin $\alpha 4$ $\beta 1$ expression was very low in both groups, suggesting that only a very small proportion of HSPCs would respond to altered fibronectin levels.

C/EBP-β and HLF expression, which had increased motif enrichment and RNA expression in CMML haematopoietic stem cells in the multiome dataset, were also included in the IMC analysis. Average expression of both TFs was no different between groups. Gene set enrichment analysis of CMML HSPCs in bulk and single cell analysis consistently found evidence of increased enrichment for genes involved in proliferation and cell division within multiple HSPC populations. However, Ki67, a well validated marker of proliferation was also no different between healthy donors and CMML patients.

Imaging mass cytometry identifies novel ligand-receptor pairs in TET2m CMML bone marrow

Bone marrow wide analysis was performed to identify particular cell types that expressed the ligands of interest discussed above. I calculated the average percentage of each cell type expressing each ligand above threshold for each individual and compared TET2m CMML patients with healthy donors (Figure 4.11A). This identified CSF1 and TGF-β as being the most widely expressed of the ligands within the bone marrow for both healthy donors and CMML patients. Interestingly, TGF-β was expressed in a greater proportion of cells within most cell types in CMML compared to healthy donors. Staining in healthy donor samples is clean, with relatively little background noise. Staining in CMML patients however is

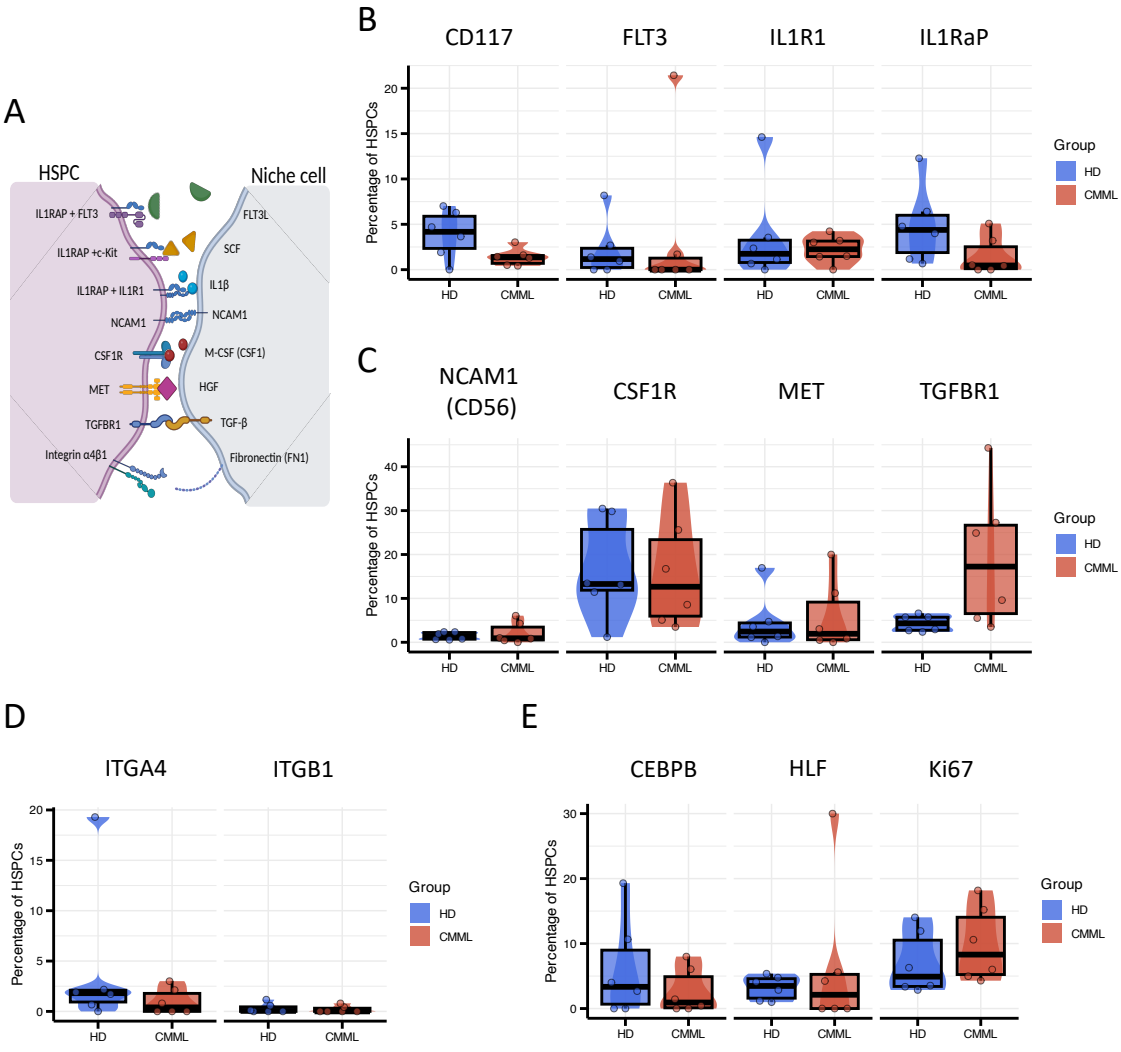


Figure 4.10: HSPCs express multiple receptors of interest but there was no difference in expression between healthy donors and TET2m CMML - (A) Schematic representation of interactions being investigated using imaging mass cytometry. Percentage of HSPCs expressing receptors above their threshold for IL1RAP and potential partners (B), receptors for NCAM1, M-CSF, HGF and TGF-β (C) and fibronectin (D). (E) Percentage of HSPCs expressing transcription factors of interest and marker of proliferation, Ki67.

almost ubiquitous, although some areas are far brighter than others (Figure 4.11B). It is likely that this difference is due to an increase in extracellular TGF-β, which is why expression appears to be increased in almost every cell type. Extracellular staining can also be seen for CSF1 (M-CSF), as well as brighter cytoplasmic staining, in both CMML and healthy donors, although it is easier to appreciate in healthy donors because of the lower cell density. M-CSF expression was particularly prominent in CSF1R positive macrophages and mesenchymal stem cells (MSC) (Figure 4.11C). However, M-CSF was also expressed on average in 20-40% of HSPCs, mature granulocytes, classical monocytes, atypical monocytes and macrophages. Proportions of cell populations expressing CD56 (NCAM1), FN1, HGF and IL-1β were generally low. NCAM1 and IL-1β staining was weak and therefore only cells expressing especially high levels of ligand were possible to confidently identify. HGF and FN1 staining however was good quality, but only a few cells were positive in each region of interest for both markers. It is likely that I have not been able to accurately identify the cell types that express HGF or FN1 with the markers I have available. I can however be confident that the cell types I have identified do not consistently express either HGF or FN1.

Classical monocytes are a source of M-CSF and TGF-β in the HSPC bone marrow niche in TET2m CMML

Spatial analysis found an increased abundance of classical monocytes and macrophages in the HSPC niche in TET2m CMML. I therefore conducted a further subgroup analysis of classical monocytes to determine which ligands may be interacting with HSPCs (Figure 4.12). Average percentage of classical monocytes expressing TGF-β was increased in CMML compared to healthy donors (23.4% vs 7.29%, p=0.02). Average percentage of classical monocytes expressing M-CSF (CSF1) was 18.9% and 22.7% in CMML and healthy donors respectively. Although there was no difference in the proportion fo M-CSF expressing monocytes in CMML, the increased proximity and abundance of monocytes in CMML increases the potential for signalling to be occurring via this axis in CMML. This can be easily appreciated by viewing

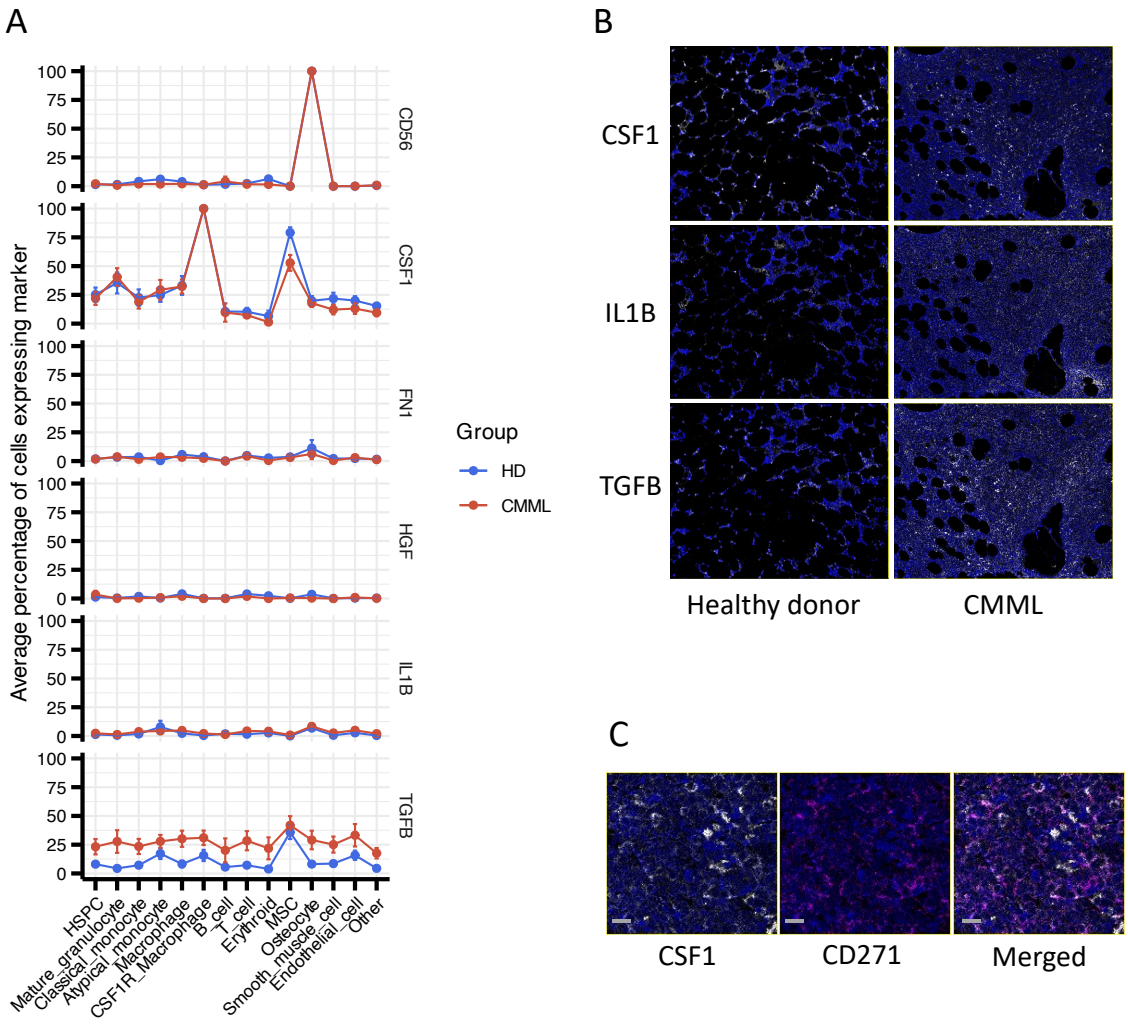


Figure 4.11: Ligand expression by different cell types within the bone marrow - (A) Average percentage of cells expressing ligands of interest calculated by individual. Error bars represent standard error of means. (B) Representative images from a healthy donor and CMML patient showing ligand expression in white and DNA staining in blue. (C) Demonstration of co-expression of CSF1 (white) and CD271 (pink), a marker of mesenchymal stem cells, in a CMML patient. Scale bar is 20μm in length.

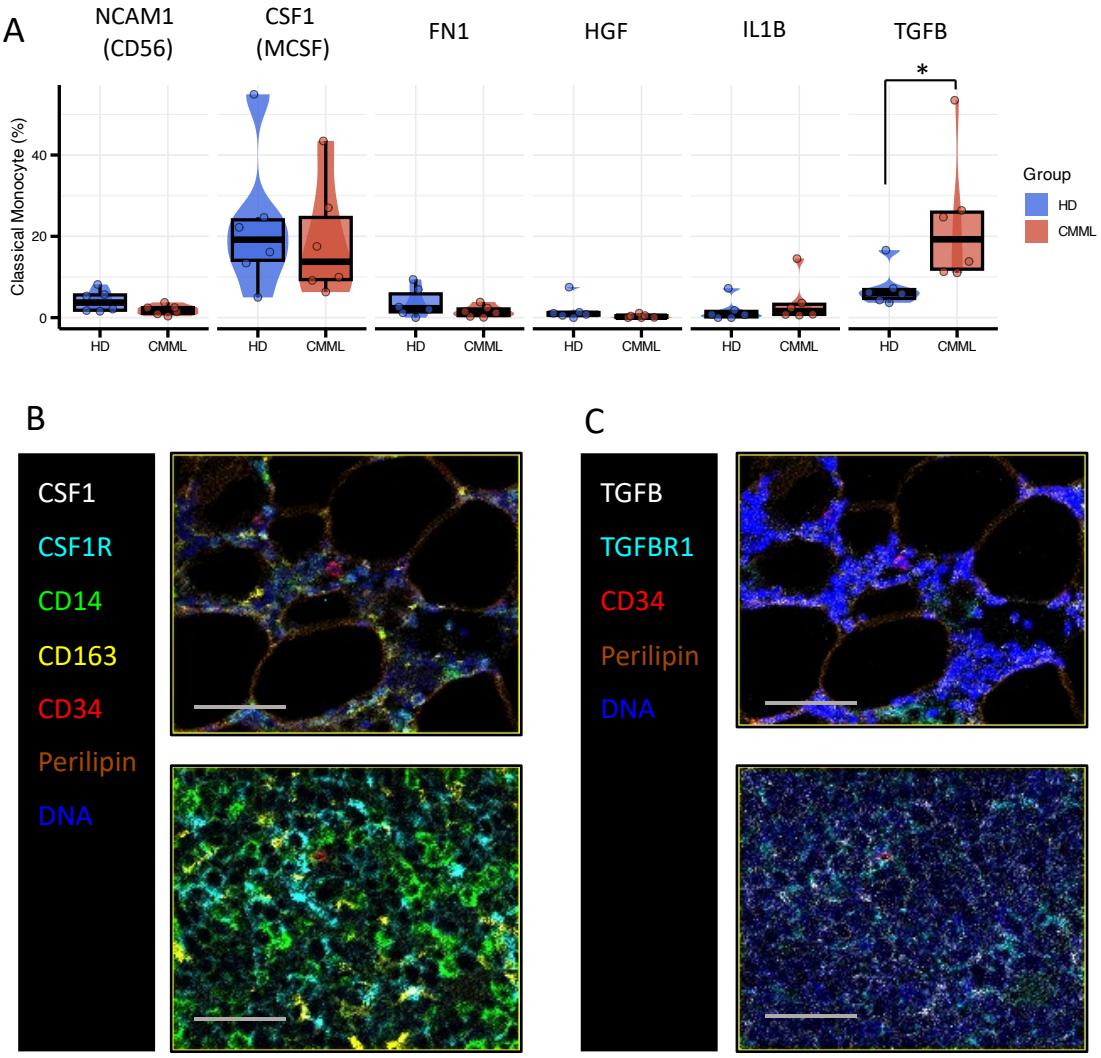


Figure 4.12: Ligand expression in classical monocytes - (A) Average percentage of classical monocytes expressing each marker in each individual. CSF1-CSF1R (B) and TGF-β-TGFBR1 (C) expression within the HSPC niche. Two representative regions containing a single HSPC were selected. Top panel is from a healthy donor and bottom panel is from a TET2m CMML patient. Scale bars are 50μm in length.

multiplex false colour staining of representative regions of bone marrow from healthy donor (upper panel) and CMML (lower panel) as shown in Figure 4.12B.

4.3 Discussion

Classical monocytes in CMML are metabolically, epigenetically and transcriptionally distinct from healthy donor monocytes[95], [110], [366]. However evidence for whether these changes are pro-inflammatory or anti-inflammatory are conflicting. I

sought to characterise the changes in TET2m CMML by using multiomic profiling of bulk sorted $CD14^+$ $CD64^+$ monocytes. As expected, TET2m CMML monocytes separated from healthy donor monocytes on principal component analysis of both gene expression and chromatin accessibility data. Pathway analysis of less accessible peaks in TET2m CMML monocytes found that these occurred at loci involved in mature myeloid cell function, indicating that TET2m CMML monocytes were less mature than healthy donor monocytes. Gene expression analysis confirmed this, where genes involved in proliferation were enriched in TET2m CMML whereas genes responsible for inflammatory signalling were enriched in healthy donors. However, when using single cell multiomic analysis this was not the case. TET2m CMML monocytes were more pro-inflammatory and less proliferative than healthy donor monocytes. This seemingly conflicting result may be explained by inter-patient heterogeneity or by $CD14^+$ monocytes being contaminated by an expanded population of immature monocytes or monocyte progenitors in CMML. Analysis of differences in cell type populations in total bone marrow mononuclear cell samples used for single cell multiomic analysis found that the latter explanation was most likely to be true. After refining cell type classification I found that GM progenitors and immature monocytes were expanded in TET2m CMML and that gene expression by GM progenitors were enriched for gene sets used in proliferation and oxidative phosphorylation when compared to monocytes that exhibited greater expression of genes involved in inflammatory response and cytokine signalling.

Ex vivo culture of CMML monocytes has found that they are more resistant to apoptosis than healthy donor monocytes via the activity of anti-apoptotic protein MCL1, which is propagated by CYTL1 autocrine signalling[95]. MCL1 has a dual function, in addition to its role in apoptosis it is responsible for regulating mitochondrial metabolism, promoting the use of oxidative phosphorylation[365]. Furthermore *Addinsell et al.* found that CMML monocytes possess increased basal respiration, maximal respiration and spare respiratory capacity, in keeping with enhanced oxidative phosphorylation[366]. Each of these studies used CD14 to positively select for monocytes, and therefore will have selected for immature

monocytes and monocyte progenitors that also express CD14. These populations are expanded in TET2m CMML. It is therefore possible that these findings reflect an inherent difference in cellular make up of samples being compared rather than specific differences in monocyte function. However, that is not to say that monocytes are unchanged in TET2m CMML. Subgroup analysis of monocytes in the single cell multiome dataset demonstrated that TET2m CMML monocytes were enriched for interferon response and inflammatory signalling pathways. Healthy donor monocytes on the other hand were enriched for NFκB targets in areas of open chromatin and gene expression of *NFKB1* was increased. These findings are suggestive of immune dysregulation, where some parts of the inflammatory pathway are active in CMML monocytes, but not necessarily via the activity of NFκB. It would be tempting to suggest that this may be a mechanism linking autoimmunity to CMML, however, the NFκB pathway relies on multiple post-translational modification steps to effect gene expression and therefore these findings only represent part of a more complex picture of immune regulation. Further work is therefore required to elucidate the significance of these findings.

The change in number and function of mature myeloid cells in TET2m CMML could also have a profound impact on HSPC function through cell-cell interactions. This is important because the cell of origin in TET2m CMML is found within the HSPC population, and therefore these interactions may support tumour growth and phenotype. I used multiple modalities to assess the changes to the number and topographical location of mature myeloid populations in TET2m CMML bone marrow. These found that classical monocytes were 10 fold increased in TET2m CMML bone marrow mononuclear cells, which was also confirmed by imaging mass cytometry in bone marrow trephine samples. Furthermore, a greater proportion of classical monocytes and macrophages were located in close proximity to HSPCs within their bone marrow niche in TET2m CMML, whereas HSPCs, mature granulocytes, T cells, erythroid cells and endothelial cells were less frequently present. These changes in cellular composition of the HSPC niche will result in a shift in the cell-cell communication HSPCs rely on for governing their function. Many tools have

been developed in recent years to try to predict changes to these communication networks using single cell RNA sequencing data[400]. Most of these tools work by creating a score for each ligand-receptor interaction between each cell type present within the data. The score is calculated by filtering gene expression matrices for potential ligands and receptors, which are identified using a curated database of ligand-receptor interactions. Statistics, such as the product of ligand and receptor expression, are then calculated to rank the likelihood of each interaction occurring. This creates a ligand-receptor interaction score matrix between each cell type, which can be visualised in many ways to identify interactions of interest for further validation. Multinichenet expands on this concept by calculating a ligand activity score, which tests the correlation between ligand expression in sender cells, with gene expression in known downstream genes of ligand-receptor interactions in receiver cells[258], [259]. This powerful additional score can be focussed on identifying ligands that are predicted to alter genes that are differentially expressed between two or more conditions. These interactions are therefore more likely to be biologically important in influencing gene regulatory circuits that are perturbed in disease.

Multinichenet analysis of single cell data, and the results from differential expression analysis of ligands and receptors in the bulk HSPC and monocyte samples, were used to identify six ligand-receptor pairs that were predicted to have altered activity in TET2m CMML. I included antibodies targeting these six ligands and 10 of their cognate receptors in the panel developed for IMC. This confirmed ligand expression of M-CSF and TGF-β in the bone marrow with cognate receptor expression on HSPCs. However, protein expression was not always concordant with what was expected from gene expression data. For instance, IL1RAP protein expression was globally low in IMC analysis, especially in CMML patients. Gene expression analysis however found that mRNA levels of *IL1RAP* were increased in TET2m CMML HSPCs. This is also at odds with previous flow cytometry data of bone marrow mononuclear cells where many CMML patients had increased IL1RAP expression, although inter-individual variability was high[360]. This discordance may be due to limitations in the performance of the antibody, changes to protein

epitope accessibility in FFPE treated samples or post translational regulation of protein expression. Each of these factors are important to consider, and that is why validation of sequencing findings is of great importance when considering their functional consequence. TGF-β ligand and receptor expression were also different to expected from gene expression data, with increased TGF-β expression detected throughout many different cell populations in TET2m CMML, suggestive of increased extracellular deposition of TGF-β. Multinichenet had predicted reduced TGFB1-TGFBR1 signalling in TET2m CMML, however IMC indicated increased presence of the TGF-β ligand and HSPC bound receptors in TET2m CMML, which most likely reflects increased signalling. This difference may be due to the way TGF-β is packaged rather than changes to the levels of TGF-β production. TGF-β is usually secreted in combination with latency-associated peptide (LAP) forming the small latent complex (SLC)[401]. This is turn can bind to latent TGF-β-binding protein-1 (LTBP-1) to form the large latent complex (LLC) that can interact with extracellular matrix proteins such as fibronectin and fibrillin[402]. LLC form deposits of latent TGF-β in the extracellular matrix, which can be unwound to release active TGF-β via the activity of integrins, reactive oxygen species, thrombospondin, proteases and changes to pH[403]. It is possible that the micro-environment in CMML prolongs the half life of SLC/LLC latent complexes resulting in increased stability and availability despite reduced production. An alternative explanation would be that TGF-β is being produced at an increased amount by a cell population that was not present in bone marrow mononuclear cells, such as macrophages, granulocytes or stromal cells, which are under-represented in ficoll-separated bone marrow aspirates. No such population however was identified using IMC, which had greater sensitivity to detect this, and therefore altered latent complex half life is an attractive explanation. Like TGF-β, M-CSF expression was present in a proportion of almost all identified cell populations by IMC, suggesting detection of either extracellular M-CSF or M-CSF signal bleeding between adjacent cells. However M-CSF expression was particularly prominent in mature myeloid cells, MSCs and the poorly defined CSF1R macrophage population. This correlates well with single

cell gene expression data where *CSF1* was found to be predominantly expressed in mature myeloid and normoblast populations. MSCs are poorly represented in bone marrow aspirate samples and as such this population was not possible to identify in the single cell data. However, erythroid cells were the least positive cell type for M-CSF in IMC data. This discrepancy can be explained by the differences in how cell populations were defined between modalities. IMC used CD235a/b (glycophorin) to identify erythroid populations, which is a late differentiation marker. However, *CSF1* gene expression was most prominent in early normoblasts, which would not have been identified by IMC. Investigation of the importance of IL-1β signalling using IMC was impossible due to poor signal from the antibody staining. However small numbers of HSPCs were found to have IL1RAP and IL1R1 expression, suggesting that altered IL-1β expression was potentially important for altering HSPC function. There are also emerging data that provides a compelling argument for IL-1β being important in the pathogenesis of MDS[404]. Therefore it was decided to also include IL-1β in our functional analysis despite the lack of validation in IMC.

The bone marrow is a particularly challenging tissue for conducting multiplexed proteomic analysis. Bone is hard and requires decalcification before sections can be cut. Even after decalcification it is challenging to maintain structural integrity when attempting sections of less than 4μm. Bone marrow sections are highly autofluorescent, making immunofluorescence microscopy in green channels particularly difficult. The bone marrow also contains a large number of different cell types that are often intermixed, requiring a large panel of markers to reliably annotate these populations. Furthermore, each of these cell types are of different size, with some extremely large cells, such as adipocytes and megakaryocytes, and others being extremely small, such as lymphocytes and erythrocytes, which makes accurate cell boundary segmentation particularly challenging. IMC was chosen for this study because it avoided the problems with auto-fluorescence and spectral overlap that impact upon sensitivity of immunofluorescence microscopy and allowed a large number of markers to be tested simultaneously. However there were many limitations to using this technique. Despite extensive optimisation several markers

were found to have poor signal, limiting the confidence in interpreting their findings. However, the greatest limitation identified in this IMC analysis was the lack of resolution. The 1μm resolution of the technique is sufficient for the analysis of many tissues, but the bone marrow is inhabited by small cells found in close proximity with one another, which made cell boundary segmentation very challenging. Despite testing many different sophisticated tools that are well validated for working with IMC data, segmentation resulted in many cells that were co-expressing markers that were inconsistent with known cell types. To mitigate for this, rather than using an unsupervised approach for annotation, I applied a stringent boolean gating strategy and excluded any cells that did not comply with these rules. This resulted in only 47% of cells being assigned a cell type. Although these cells were evenly distributed throughout regions of interest many unidentified cells were found within the HSPC niche and therefore it is possible that important changes in cell type distribution in this niche were missed or falsely represented. Furthermore, many cells that were annotated a cell type could also be found to be expressing markers of adjacent cells, confounding the analysis of ligand and receptor expression. Recent advances in multiplexed immunofluorescence and spatial transcriptomics have allowed the detailed spatial analysis of bone marrow. The limitations I encountered using IMC are therefore possible to overcome and their increased power to detect a greater number of cell types allows for the integration of spatial data with single cell transcriptomics to enhance cell-cell interaction analysis. It is likely that these techniques will be fundamental to how these tissues are investigated in the future.

4.3.1 Conclusion

Evidence of monocyte dysfunction in CMML has been conflicting. I demonstrate that bulk CD14$^+$ monocyte populations include a greater proportion of immature monocyte progenitors in CMML leading to a more proliferative and less inflammatory phenotype. However single cell analysis of purified mature monocytes found that TET2m CMML monocytes had increased inflammatory cytokine expression but with down regulation of the canonical NFκB pathway. Monocyte populations were

markedly expanded in TET2m CMML bone marrow with a trend for increase in macrophages. Monocytes and macrophages were also positioned in closer proximity to HSPCs in TET2m CMML, implicating them as potentially important extrinsic regulators of HSPC function. M-CSF, TGF-β and IL-1β were selected as ligands that had good evidence to support their altered activity in TET2m CMML and therefore the functional implications of these molecules to TET2m CMML HSPCs is the focus of the next chapter.

5

Ligand-receptor interactions can be targeted to inhibit cancer cell growth and function in TET2m CMML

Contents

5.1 Introduction 160
5.1.1 Investigating the extrinsic regulators of haematopoiesis 160
5.1.2 M-CSF and CSF1R in haematopoiesis 161
5.1.3 Targeting M-CSF and CSF1R in cancer 162
5.1.4 TGF-β signalling in haematopoiesis 164
5.1.5 IL-1β signalling in haematopoiesis 165
5.1.6 Summary and plan for investigation 166
5.1.7 Aims 166
5.2 Results 167
5.2.1 TET2m CMML classical monocytes express CSF1R and NCAM1 167
5.2.2 CSF1R inhibition with emactuzumab leads to a dose dependent reduction in monocyte growth in the presence of M-CSF 168
5.2.3 Assays to test for dysregulation of HSPC response to extrinsic cytokine signalling in TET2m CMML 170
5.2.4 IL-1β exerts a proliferative effect on HSPC culture and promotes myelomonocytic differentiation in TET2m CMML and healthy donors 173
5.2.5 TGF-β causes dose dependent growth inhibition, suppression of myelomonocytic differentiation and enhanced erythropoiesis in TET2m CMML and healthy donor HSPCs 173

5.2.6 Emactuzumab inhibits proliferation and monocyte differentiation of HSPCs in M-CSF responsive TET2m CMML patients 177
5.3 Discussion 181
5.4 Conclusion 186

5.1 Introduction

5.1.1 Investigating the extrinsic regulators of haematopoiesis

Colony forming assays were initially developed in the 1960s after it was discovered that haematopoietic bone marrow derived cells were capable of forming colonies when grown in semi-solid culture conditions[405]. These assays facilitated the identification of cytokines that are fundamental to the regulation of haematopoiesis. They have also, along with *in vivo* transplantation assays, allowed for the identification of the diverse spectrum of multipotent and oligopotent subtypes of HSPC. HSCs are largely quiescent throughout life, with multiple redundant mechanisms protecting them from leaving this state[406]. However, when HSCs are triggered to enter the cell cycle they then undergo either symmetrical or asymmetrical cell division. Symmetrical division results in two daughter cells identical to the parent, a process of self renewal, and asymmetrical division results in two different daughter cells, one which preserves multipotency and another that becomes more restricted in the lineage of its progeny, a process known as differentiation. This complex and dynamic process is coordinated by instructive and permissive cytokine signals that promote specific HSPC subtype proliferation and survival respectively[407]. HSCs possess many different cytokine receptors that are often expressed at very low levels and can be difficult to detect. However, signalling via these receptors can result in the creation of a positive feedback loop by promoting increased cognate receptor expression and in doing so support lineage differentiation. M-CSF/CSF1R signalling, as discussed below, is an example of such cytokine signalling.

Extrinsic regulation is not only important for influencing haematopoietic stem cell function, but is also instructive for lineage commitment and proliferation of

more mature haematopoietic cell types. This adaptive regulation of haematopoiesis, as discussed in the case of emergency myelopoiesis in previous chapters, is crucial for responding to physiological stressors at an organism level such as infection or haemorrhage. Monocyte production is therefore one of these processes that can be extrinsically regulated and therefore targeting ligand-receptor interactions that may suppress monocyte production may be effective in modifying CMML phenotype. This is especially relevant because monocytes, as demonstrated in the previous chapter, have an increased presence within the HSPC niche in TET2m CMML.

5.1.2 M-CSF and CSF1R in haematopoiesis

M-CSF (or CSF1) was the first colony stimulating factor to be purified[408] and was found to promote macrophage colony formation in haematopoietic progenitors and proliferation of peritoneal macrophages *in vitro*[409]. M-CSF exists as multiple different isoforms. It can be a membrane bound glycoprotein or a secreted glycoprotein or proteoglycan. Both the secreted and membrane bound isoforms are important in regulating haematopoiesis[410]. CSF1R, a class III receptor tyrosine kinase of the platelet-derived growth factor family, functions as the cognate receptor for M-CSF and another cytokine, IL-34, which has a separate spatiotemoporal expression profile to M-CSF[411], [412]. After a ligand binds to CSF1R, the ligand-receptor complex undergoes oligomerisation, conformational change to expose the catalytic domain and signal transduction through two waves of tyrosine phosphorylation. CSF1R then becomes ubiquitinated by the E3 ubiquitin-protein ligase CBL, internalised and degraded[413], [414]. Signal transduction is triggered through multiple pathways, some of which are cell type specific, resulting in its pleiotropic effects. Phosphoinositide 3-kinase (PI3K)/AKT signalling promotes macrophage survival and chemotaxis, whereas Syc-family kinase signalling and ERK1/2 supports monocyte/macrophage differentiation and macrophage proliferation[414]. In addition to cell-surface bound CSF1R, a tiny fraction of CSF1R is trafficked to the nucleus, where it can bind directly with DNA and alter YY1 and ELK transcription factor function upon M-CSF stimulation[415].

CSF1R is expressed at low levels on HSCs[416]. Stimulation of HSCs with M-CSF induces PU.1 expression, the master regulator of myelomonocytic differentiation. PU.1 in turn binds to the promoter and fms-intronic regulatory element (FIRE) enhancer of CSF1R, creating a feed forward loop supporting myeloid differentiation[417]. CSF1R expression increases throughout differentiation, with maximal expression found in macrophages and other mononuclear phagocytic cells. CSF1R signalling is important for maintaining tissue macrophage populations, promoting proliferation, chemotaxis and survival, and has also been implicated in regulating macrophage polarisation, supporting the development of M2-type homeostatic macrophages[418].

CSF1R is critical during the embryonic development and maintenance of many other tissues including osteoclasts in the bone[419], microglia and neural progenitor cells in the brain[420], oocytes and pre-implantation embryos[421], renal proximal tubule epithelial cells, small intestinal Paneth cells and colonic epithelial cells[414], [422]. CSF1R knockout therefore results in many abnormalities, such as small stature, failure of tooth eruption, osteopetrosis, reduced monocytes and lymphocytes with increased granulocytes in the peripheral blood, reduced tissue macrophages, prolonged menstrual cycle and failure of mammary gland development[423]. In humans, inactivating mutations of CSF1R can result in adult-onset leukoencephalopathy and inappropriate expression of CSF1R has been implicated in the development of leukaemia, lymphoma and disease progression in solid tumours[424], [425].

5.1.3 Targeting M-CSF and CSF1R in cancer

M-CSF/CSF1R signalling is a currently unexplored axis for treating CMML, but there are some data that support its altered activity in CMML. Plasma levels of M-CSF are increased 2.6-fold in CMML compared to age matched healthy controls[109]. CSF1R mRNA expression is increased in peripheral blood mononuclear cells[112]. CSF1R activity however may be influenced by additional exogenous factors, such as α-defensins, which have been reported to be released by immature granulocytes

and inhibit CSF1R activity in CMML monocytes[113]. CSF1R is also present on CMML blasts and HSPCs[17], [360], [426]. Furthermore, pre-clinical data using pacritinib, a JAK1, IRAK1 and CSF1R inhibitor, and early phase human trials using NMS-03592088, a novel FLT3, KIT and CSF1R inhibitor, have shown some promising early signs of efficacy in CMML[99], [427]. However, whether this is due to their specific anti-CSF1R activity is unknown. It therefore remains to be seen whether M-CSF/CSF1R signalling is important for driving CMML tumour development and whether specifically targeting this axis would be clinically effective.

Targeting CSF1R has however been extensively studied in AML, with CSF1R expression found on some patient's blasts[428], and CSF1R signalling being required for xenograft transplantation of primary inv(16) AML cells[429]. Furthermore, the constitutively active RBM5::CSF1R fusion protein has been discovered as a rare driver of acute megakaryoblastic leukaemia[430]. However, evidence for CSF1R as a direct driver of AML are rare. Far more commonplace are data demonstrating changes in the tumour micro-environment in AML, where CSF1R expression can be found on haematopoietic cells that support AML cell growth. Inhibiting CSF1R activity through siRNA or small molecule inhibition negatively impacts AML cell growth by reducing paracrine signalling[346]. This phenomenon is also found in solid organ tumours, where CSF1R expression on tumour associated macrophages has been targeted to inhibit M2 immunosuppressive macrophage polarisation. Multiple inhibitors for CSF1R have been developed including emactuzumab, a humanised IgG monoclonal antibody that binds irreversibly to CSF1R, preventing dimerisation and thereby inhibiting downstream signalling[431]. Emactuzumab has been licensed for use in tenosynovial giant-cell tumours[432] and has also demonstrated promising pre-clinical activity in depleting tumour associated macrophages and switching macrophage polarisation to create a less immunosuppressive environment in multiple solid organ malignancies[431]. We set up a collaboration with its manufacturers, SynOx Therapeutics, to investigate its utility in targeting TET2m CMML.

5.1.4 TGF-β signalling in haematopoiesis

Transforming growth factor-β (TGF-β) is part of a large family of secreted polypeptide growth factors that includes activins and bone morphogenetic proteins. TGF-β is important for the regulation of homeostasis and regeneration of many different tissues, including those involved in haematopoiesis. TGF-β is usually secreted in an inactive form (LAP), that becomes amalgamated into a large latent complex (LLC) with LTBP-1. LLC associates with extracellular matrix proteins to create a pool of latent TGF-β[402]. LLC is unwound to release active TGF-β via the activity of integrins and other micro-environmental factors[403]. Active TGF-β dimers can bind with different avidity and specificity to a diverse spectrum of type I and type II serine/threonine kinase receptors with varied functional consequences depending on the cellular context. However, TGF-β has greatest specificity to heteromeric TGF-β receptor-I (TβR-I) and TGF-β receptor-II (TβR-II) complex. Formation of this ligand-receptor complex leads to phosphorylation of TβR-I, which in turn activates SMAD proteins and leads to signal transduction to the nucleus[433].

In haematopoiesis, TGF-β signalling promotes HSC quiescence by inhibiting enzymes that are crucial within the cell cycle[379], [434]. However, the effects of TGF-β are not uniform throughout the HSC population, with proliferation of myeloid-biased HSC being stimulated and lymphoid-biased HSC being inhibited[435]. TGF-β inhibits erythroid progenitor proliferation and induces rapid terminal differentiation, overall exerting a negative impact on erythropoiesis[436]. This effect, and the inhibitory impact of other TGF-β superfamily ligands, has recently been targeted using activin receptor-IIA ligand traps to enhance erythropoiesis in thalassaemia and MDS[80], [437], [438].

The effect of TGF-β on myelopoiesis is complex, with some reports suggesting it is supportive and some inhibitory. It is likely that the interplay between cytokines and TGF-β are important for determining its effect. For instance, *Tgfb1* knockout mice develop enhanced myelopoiesis and a fatal inflammatory disorder, indicating TGF-β is an important negative regulator of myelopoiesis[380]. However, TGF-β acts synergistically with GM-CSF to promote myeloid progenitor proliferation and

M-CSF to promote macrophage proliferation[439], [440]. It is also essential for dendritic cell differentiation[441]. This nuanced, cell type specific effect of TGF-β makes it hard to make generalisations about its overall effect on myelopoiesis.

5.1.5 IL-1β signalling in haematopoiesis

IL-1β is a pro-inflammatory cytokine produced almost ubiquitously by cells in response to an inflammatory stimulus, although myeloid innate immune cells exhibit particularly strong expression[442]. IL-1β is translated as a pro-protein that requires proteolytic cleavage by either a caspase-dependent inflammasome or an alternative enzyme to create its active isoform. Active IL-1β is then secreted from the cell through temporary membrane permeabilisation[443]. Pyroptosis, a caspase-1 dependent form of programmed cell death, has also been identified as a source of active IL-1β, where it is released as the cell dies[444]. IL-1β binds to the transmembrane receptor IL-1R1 inducing conformational change and recruitment of IL-1 receptor accessory protein (IL-1RAP). MYD88 then binds to the C-terminal Toll and IL-1 R-like (TIR) domains of this trimeric complex to form the Myddosome complex that can phosphorylate IRAK proteins. pIRAK proteins prompt further signal transduction via tumour-necrosis factor-associated factor 6 (TRAF6), culminating in activation of NF-κB[445]. NF-κB is an important mediator of the inflammatory response and promotes the creation of cytokines and chemokines by the activated cell, which recruits leukocytes to the area to combat infection, control tissue damage and amplify the inflammatory signal[446].

This amplified inflammatory signal can be sensed throughout the body, including bone marrow resident HSCs. This signal promotes emergency myelopoiesis, as discussed in previous chapters, and IL-1β is a direct mediator of this. IL-1β induces PU.1 activity in a subset of HSCs, accelerating their myeloid differentiation. It also suppresses self-renewal, through promoting HSC quiescence, maintaining homeostasis within the HSC pool during inflammation[447], [448]. It is therefore not all that surprising that IL-1β has been implicated as an important driver of myeloid disease, including those driven by TET2 loss of function. IL-1β, along with IL-6,

and TNF-α, are increased in murine models of TET2m clonal haematopoiesis[41], [42]. IL-1β has also been implicated in supporting HSPC expansion and self renewal in TET2m clonal haematopoiesis during aging[184], [382]. Chronic exposure to IL-1β in *Tet2* knockout mice leads to monocyte expansion and development of a CMML-like phenotype. Furthermore, IL-1β enhances methylation of target sites of transcription factors that regulate terminal haematopoietic cell differentiation, including C/EBP-β and BACH2, in *Tet2* KO LSK cells[449]. IL-1β has also been identified as an important mediator of MDS pathogenesis and is emerging as a novel therapeutic target in MDS and AML[383], [450].

5.1.6 Summary and plan for investigation

Haematopoiesis is controlled by the complex interplay of intrinsic and extrinsic signals at every level throughout differentiation. However, the altered regulation of HSPCs, which sit at the apex of this process, is of most relevance to TET2m CMML pathogenesis. By screening for aberrant expression and activity of extrinsic regulators of haematopoiesis I identified M-CSF, TGF-β and IL-1β as ligands to investigate further in TET2m CMML. Each of these ligands is important for promoting myelopoiesis and therefore may induce enhanced activity in epigenetically primed TET2m CMML HSPCs. To investigate this hypothesis I created multiple *in vitro* primary culture assays to assess the effect of these ligands on HSPC function and in the case of M-CSF on monocyte growth.

5.1.7 Aims

The specific aims for this section of my thesis are:

1. Investigate the effect of M-CSF supplementation and CSF1R inhibition on TET2m CMML monocyte growth dynamics

2. Investigate the impact of M-CSF, TGF-β and IL-1β on TET2m CMML HSPC proliferation and survival

3. Identify whether targeting these axes would provide potential benefit to TET2m CMML patients

5.2 Results

5.2.1 TET2m CMML classical monocytes express CSF1R and NCAM1

Imaging mass cytometry of the bone marrow found that the number of classical monocytes are greatly increased in the bone marrow of patients with TET2m CMML and that they reside in close proximity to HSPCs, making it likely that they can influence their function through cell-cell interactions. Therefore, identifying novel treatments to target classical monocytes in TET2m CMML may have a positive impact on patient prognosis, or act synergistically with a treatment that targets the HSPCs. I attempted to screen cytodelics whole blood stabiliser preserved peripheral blood for NCAM1 and CSF1R expression, but antibodies targeting CSF1R performed poorly (data not presented). CSF1R is known to be difficult to assay because it is internalised after blood is collected unless cells are kept at low temperature. Expression is also diminished in thawed cells, precluding its accurate measurement in cryopreserved peripheral blood mononuclear cells. CSF1R expression however can be recovered by an overnight culture in RPMI 1640 media with 10% FBS at 37°C[451]. I therefore cultured lymphocyte depleted peripheral blood mononuclear cells for 3 to 6 hours and measured CSF1R and NCAM1 expression in $CD14^+$ $CD16^-$ classical monocytes (Figure 5.1A). PE-conjugated emactuzumab was used to detect CSF1R expression. The performance of emactuzumab was evaluated by co-staining samples with a commercial antibody towards CSF1R. Correlation was good between antibodies ($r=0.74$, $p=0.02$) and <10% lymphocytes were bound by emactuzumab, confirming specificity. Using emactuzumab-PE, over 90% of classical monocytes from all four TET2m CMML patients tested were CSF1R positive compared to 54% on average in healthy donors (Figure 5.1C and D). NCAM1 was expressed in a very high proportion of classical monocytes in three TET2m CMML patients, but not at all in one patient or any

of the healthy donor controls. Neither CSF1R or NCAM1 has high expression in healthy donors, and therefore targeting these molecules with immunotherapy may be a successful strategy for removing malignant monocytes in CMML.

5.2.2 CSF1R inhibition with emactuzumab leads to a dose dependent reduction in monocyte growth in the presence of M-CSF

M-CSF is an important cytokine for directing monocyte maturation, proliferation and survival. CSF1R is usually expressed at high levels on atypical monocytes ($CD14^{weak}$ $CD16^{+}$) and to a lesser extent on classical monocytes[451]. CSF1R inhibition therefore has a disproportionate effect on atypical monocytes, leading to cell death in vitro and in vivo[431]. However in TET2m CMML it may have a more exaggerated effect on classical monocytes because they display increased expression of CSF1R. To test this hypothesis I isolated monocytes by negative selection using the same technique outlined above. Following a brief culture in RPMI1640 supplemented with 10%FBS, monocytes were treated with a range of doses of emactuzumab, from 0.001 to 10μg/ml. After 24 hours in culture, viable monocyte cell number was assessed by CellTiter Glo and monocyte cell death was assayed by CellTox Green (Figure 5.2A). Monocyte cell number in healthy donors and TET2m CMML patients was reduced by emactuzumab in a dose dependent manner. Although there was no statistical difference between dose response curves, there was a trend towards healthy donor monocytes being more sensitive to emactuzumab than TET2m CMML patient monocytes (Figure 5.2B). CellTox Green however was not increased, suggesting that this inhibition in cell number was secondary to reduced proliferation rather than increased cell death (Figure 5.2C). Previous unpublished pre-clinical work on emactuzumab in human monocyte derived macrophages found that emactuzumab efficacy increases with longer culture durations. It also found that concomitant exposure to M-CSF or GM-CSF altered sensitivity to emactuzumab. This is potentially of great relevance to CMML, where GM-CSF hypersensitivity has previously been reported[97]. Therefore, the assay was repeated using a 72 hour culture, with supplementation with either 100ng/ml of M-CSF or GM-CSF.

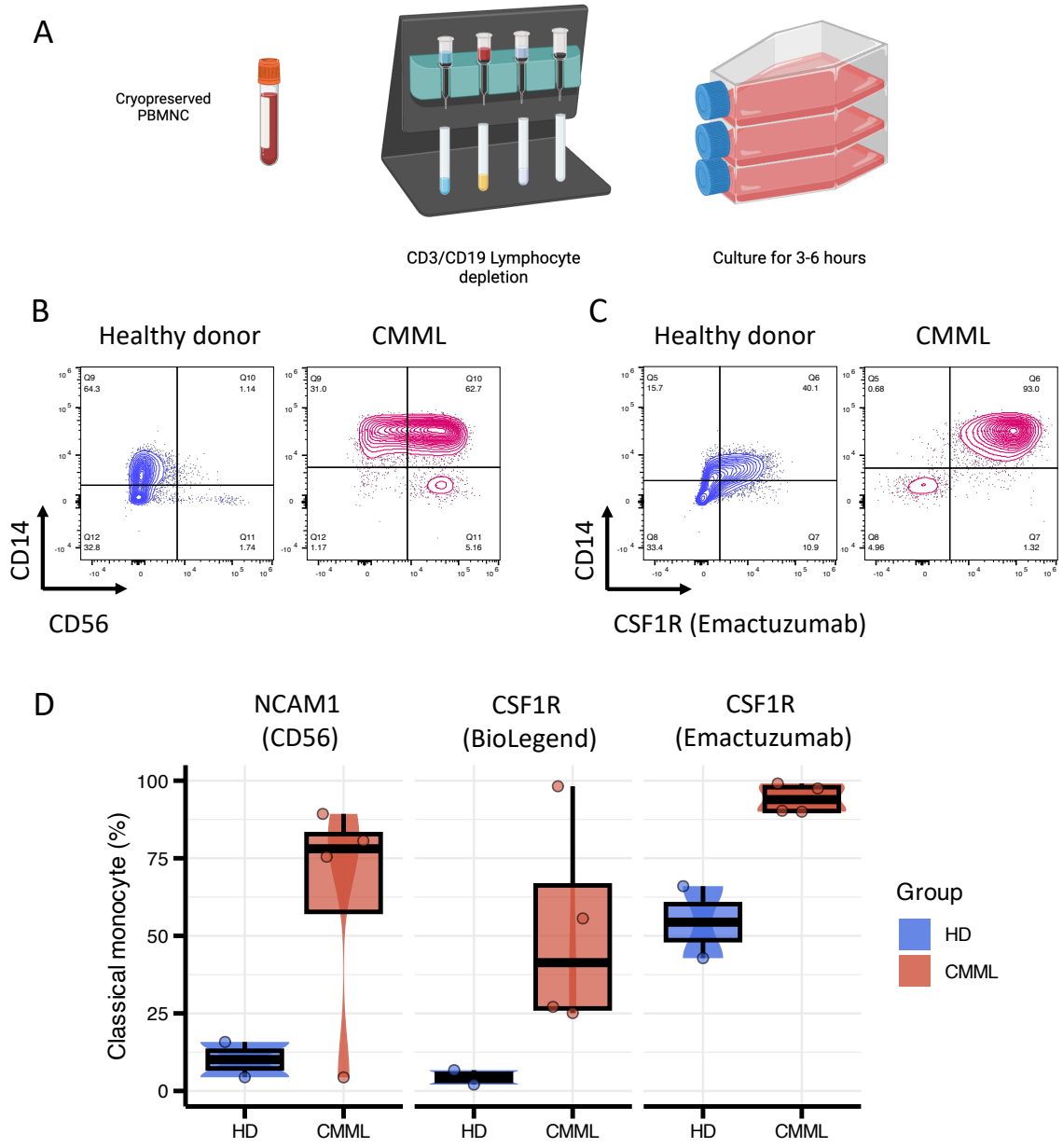


Figure 5.1: NCAM1 and CSF1R are expressed by classical monocytes in TET2m CMML - (A) Visual summary of monocyte isolation by magnetic column lymphocyte depletion of cryopreserved peripheral blood mononuclear cells and brief culture in RPMI 1640 with 10% FBS prior to assay by flow cytometry. Representative flow cytometry plots of NCAM1 (CD56) expression (B) and CSF1R expression (C) in $CD14^{+}$ classical monocytes. (D) Summary of NCAM1 expression and CSF1R expression by a commercial and PE conjugated therapeutic antibody. Differences between groups were not significant by Wilcoxon test.

After 72 hours of treatment with emactuzumab, monocyte growth inhibition was more profound in M-CSF supplemented monocytes than after 24 hours with a dose dependent increase in inhibition (Figure 5.2D). However, there was no difference between TET2m CMML patient monocytes and healthy donor monocytes. Culture with GM-CSF supplementation diminished the effect of emactuzumab (Figure 5.2E). There was no statistical difference between TET2m CMML monocyte and healthy donor monocyte dose response curves, although given that monocyte counts are 5-10 fold increased in CMML, a similar effect on proliferation will have a far more profound effect in CMML patients than in healthy donors.

5.2.3 Assays to test for dysregulation of HSPC response to extrinsic cytokine signalling in TET2m CMML

Two assays were designed to test for changes in HSPC function in response to stimulation by exogenous ligand or inhibition of these interactions (Figure 5.3A). Changes of particular interest were alterations to differentiation bias, proliferation, cell death or quiescence. HSPC proliferation and differentiation were assessed in the proliferation assay where HSPCs were cultured in serum free conditions in the presence of the specific ligand of interest along with a minimal cocktail of necessary cytokines to support HSPC survival. After 8 days, cell number and cell type were assessed by flow cytometry. HSPC survival was assessed by culturing HSPCs in serum free conditions with only the specific ligand of interest supplemented to the media. After 48 hours, a complete cocktail of haematopoietic cytokines were added and HSPCs were cultured for 11 further days. This survival assay assesses how candidate interactions alter HSPC apoptosis and quiescence by comparing total cell numbers and infer HSPC subpopulation survival by measuring the specific cell types that proliferate after the complete cocktail of cytokines is added. Output from cultures was assayed by flow cytometry, with cells being stained within their culture plates to minimise cell loss. Counting beads were also added after staining to ensure accuracy of cell counts across different samples and plates.

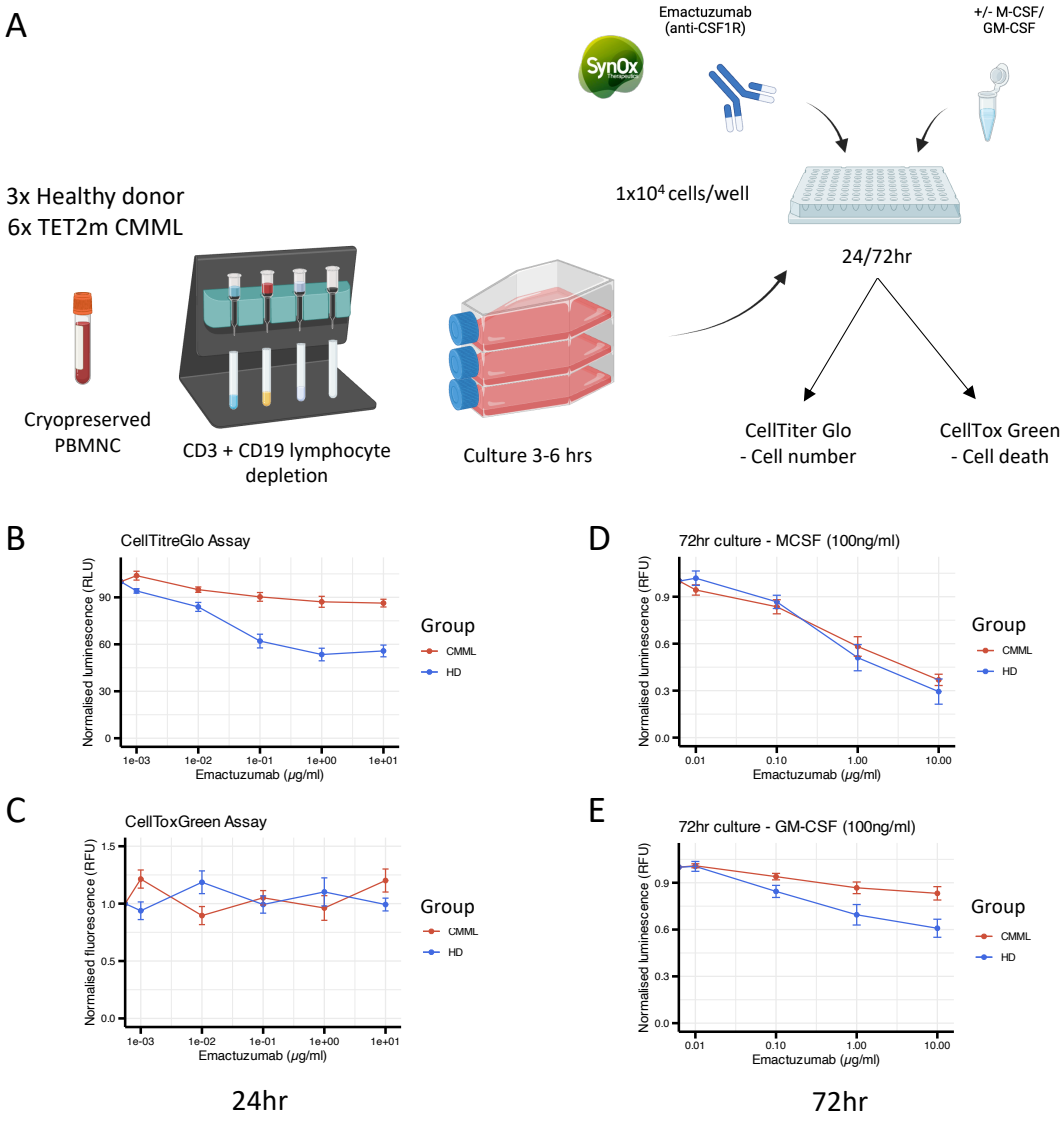


Figure 5.2: Emactuzumab inhibits monocyte proliferation in a dose dependent manner in TET2m CMML and healthy donor monocytes - (A) Visual summary of monocyte isolation by negative selection, followed by a brief culture in RPMI 1640 with 10% FBS to restore CSF1R expression. Monocytes were then counted and transferred to plates where they were treated with a range of emactuzumab doses. After 24 hours of culture monocytes were assessed for cell number by CellTiter Glo (B) or cell death by CellTox Green (C). Luminescence or fluorescence was normalised to vehicle control wells for each individual. Monocytes cultured for 72 hours were supplemented with either M-CSF (D) or GM-CSF (E) and assayed by CellTiter Glo.

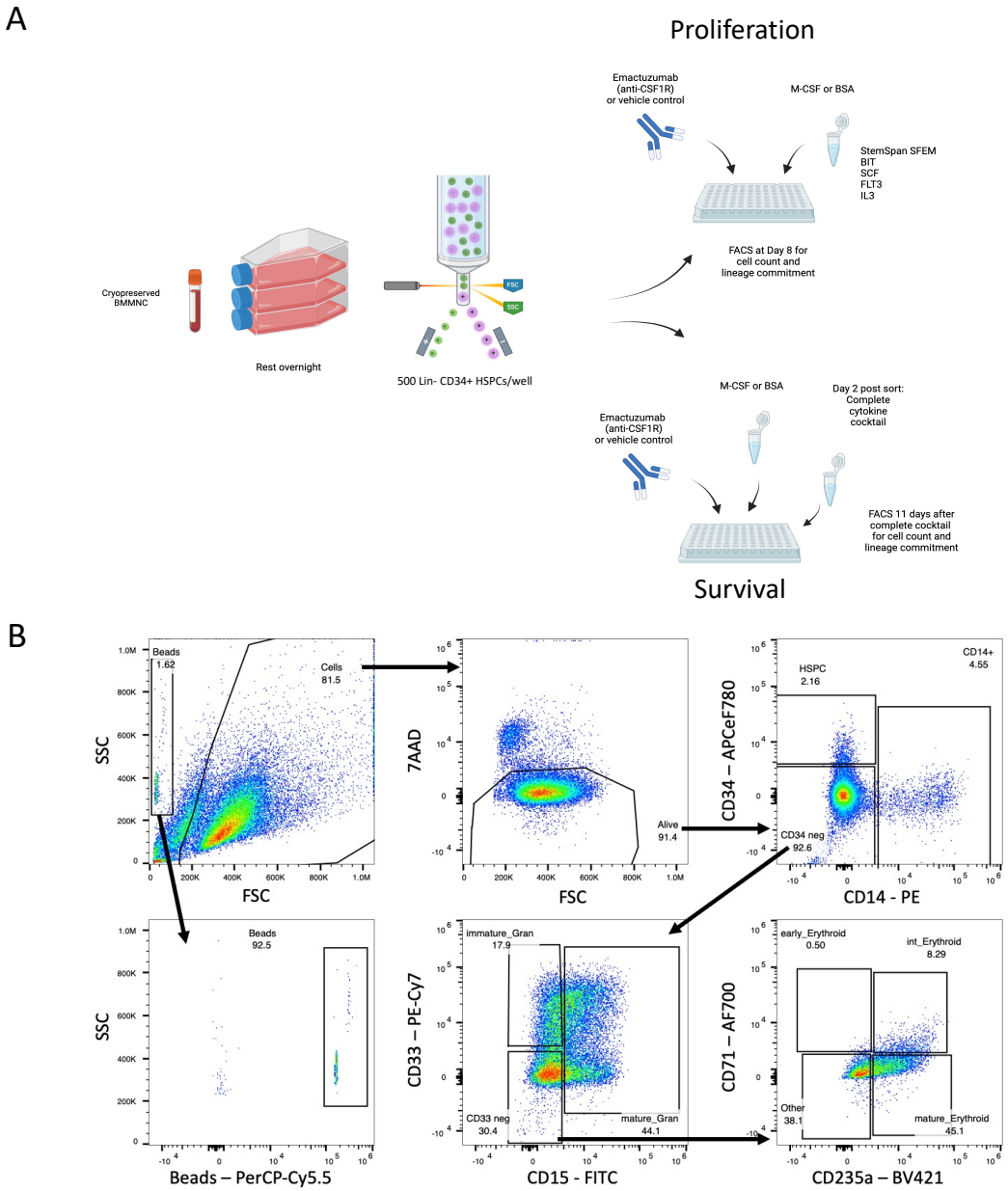


Figure 5.3: Primary HSPC culture to assess proliferation and survival in response to stimulation or inhibition of ligand-receptor interactions - (A) Summary of HSPC proliferation and survival assays. Cryopreserved HSPCs were thawed and rested overnight in serum free media and a minimal cocktail of supportive cytokines. Lineage$^-$ CD34$^+$ HSPCs were sorted by FACS into 96 well plates with the cytokine of interest and a minimal cocktail of supportive cytokines. Assays displayed here are for investigation of M-CSF-CSF1R interactions, but similar assays were designed for investigating TGF-β and IL-1β. (B) Flow cytometry was used to count alive cells and assess HSPC differentiation to erythroid and myeloid cell types.

5.2.4 IL-1β exerts a proliferative effect on HSPC culture and promotes myelomonocytic differentiation in TET2m CMML and healthy donors

IL-1β exerts a proliferative effect on HSCs and promotes myeloid differentiation under chronic exposure[447]. To investigate the impact of IL-1β on TET2m CMML, HSPCs were cultured with a low dose of supportive cytokines (10ng/ml FLT3L, SCF and IL-3) and either BSA or 0.1 to 10ng/ml IL-1β for 8 days. This found a dose dependent proliferative effect of IL-1β on both healthy donor and TET2m CMML HSPCs. Overall, when comparing BSA with 10ng/ml IL-1β, total cell numbers were increased on average 1.2 fold in both healthy donors and TET2m CMML (mean viable cells 93608 vs 114923, p=0.02; and 145365 vs 179678, p=0.02; respectively). However, this effect was heterogeneous and there was no difference between disease groups (Figure 5.4A and B). 10ng/ml IL-1β promotes both immature granulocyte (104065 vs 136036, p=0.02) and monocyte expansion in TET2m CMML (2492 vs 3968, p=0.04), but promotes immature granulocyte expansion (46734 vs 60286, p=0.01) and monocyte count reduction (1102 vs 451, p=0.04) in healthy donors. IL-1β inhibition would therefore be predicted to have a suppressive effect on myelomonocytic cell production in TET2m CMML, especially as IL-1β expression has been found to be increased in TET2m CH and myeloid malignancies. However, the proliferative response to IL-1β was heterogeneous in both TET2m CMML patients and healthy donors[383], [449], [450]. Therefore, further work is required to identify whether this may be a targetable axis in CMML and in whom it would provide the greatest benefit.

5.2.5 TGF-β causes dose dependent growth inhibition, suppression of myelomonocytic differentiation and enhanced erythropoiesis in TET2m CMML and healthy donor HSPCs

TGF-β promotes HSC quiescence and some studies suggest it negatively regulates early myeloid differentiation[379], [380]. TGF-β signalling was predicted to be downregulated in TET2m CMML by multinichenet but bone marrow IMC found

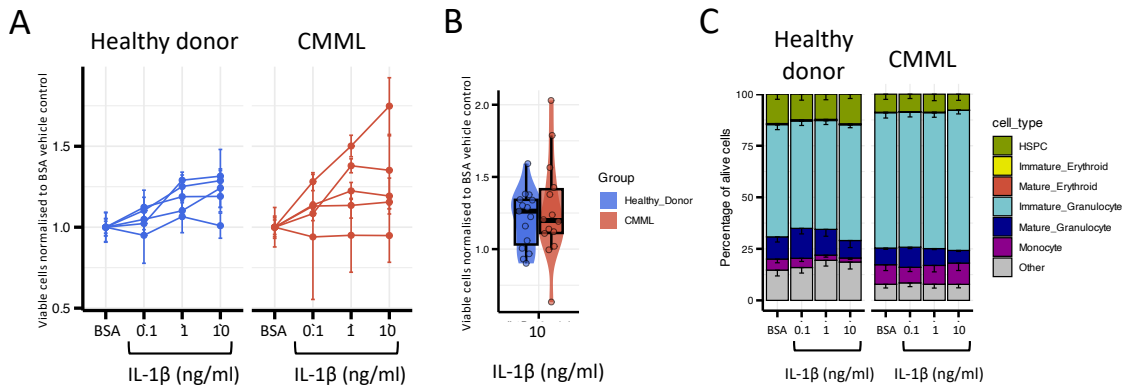


Figure 5.4: IL-1β has a proliferative effect on HSPC culture and promotes myelomonocytic differentiation in TET2m CMML and healthy donors - (A) Proliferation assay output. HSPCs were cultured for 8 days with a supportive cocktail of cytokines and either BSA or 0.1 to 10ng/ml IL-1β. Viable cell counts were normalised to BSA cultures for each individual and split by disease group. (B) Box and whisker plot of 10ng/ml IL-1β supplemented cultures. Each point represents an independent culture replicate. (C) Average percentage of each cell type by individual. Error bars in all plots represent standard error of means.

that TGF-β ligand expression was no different between TET2m CMML patients and healthy donors. This could indicate that TET2m CMML HSPCs are less sensitive to TGF-β. To test this theory, I modified the HSPC proliferation assay so that it could detect the inhibitory effect of increasing doses of TGF-β when added to the complete, proliferative cocktail of cytokines used in the HSPC survival assay. HSPC growth was inhibited in a dose dependent manner in all individuals tested (Figure 5.5A). However there was heterogeneity between individuals, with both healthy donor and TET2m CMML patients splitting into high and low responder groups. Given the limited numbers of individuals tested however, it is not possible to discern whether these are groups that would be found in the wider population. Compared to BSA supplemented control cultures, viable cells after culture with 10ng/ml TGF-β supplementation were lower in both healthy donors and TET2m CMML patients (408590 vs 216409, p=0.02; 530250 vs 183395, p=3.8x10^{-4}). Inhibition was cell type specific, with significant reductions in HSPCs (13853 vs 3848, p=5.8x10^{-7} in healthy donors; 19346 vs 5872, p=2.4x10^{-5} in TET2m CMML), immature granulocytes (220151 vs 67435, p=.0003 in healthy donors; 314831 vs 50775, p=8.1x10^{-5} in TET2m CMML) and monocytes (21325 vs 5624, p=0.002 in healthy donors; 52435

vs 29312, p=0.003 in TET2m CMML), but increases in intermediate erythroid cells (941 vs 3795, p=4.4x10^{-5}; 263 vs 3041, p=3.4x10^{-5} in TET2m CMML) in both healthy donors and TET2m CMML patients. However, when comparing the total number of cells, the inhibition in TET2m CMML patients was significantly greater than in healthy donors (Figure 5.5B). TGF-β also induced a dose dependent increase in immature and mature erythroid cells indicating its role in promoting terminal erythropoiesis (Figure 5.5C).

The effect of TGF-β on HSPC survival was assessed by culturing HSPCs in serum free culture media with 10ng/ml SCF, 10ng/ml FLT3L and either BSA or 0.01 to 10ng/ml TGF-β for 48 hours followed by the addition of a complete proliferative cocktail of cytokines and cultured for a further 11 days. Cytokines were replenished every 72 hours. SCF and FLT3L were used alongside TGF-β to prevent excess HSPC death during the initial 48 hour culture. HSPC survival in response to TGF-β was highly heterogeneous in both healthy donors and TET2m CMML, with no discernible pattern found in the limited number of individuals assayed (Figure 5.5D). Overall, both healthy donors and TET2m CMML patients had reduced total viable cells after culture with 10ng/ml TGF-β compared to BSA (315768 vs 185175, p=4.1x10^{-5} for healthy donors; 267680 vs 141305, p=0.001 for TET2m CMML), with an altered pattern of cell population changes to those seen in the proliferation assay. HSPC numbers were reduced (16075 vs 4139, p=4.1x10^{-5} in healthy donors; 22282 vs 10699, p=0.05 in TET2m CMML), as were immature granulocytes (159281 vs 98868, p=0.006 in healthy donors; 100928 vs 41962, p=0.0002 in TET2m CMML) but no change was seen in monocytes(81210 vs 63652, p=0.6 in healthy donors; 89160 vs 60191, p=0.3 in TET2m CMML). Erythroid cells were only rarely detected at all dose levels of TGF-β, indicating that TGF-β's role in regulating terminal erythroid differentiation is not through the survival of erythroid progenitors but in the supporting their proliferation during differentiation (Figure 5.5E). There was therefore no evidence that TET2m CMML HSPCs were resistant to the suppressive effects of TGF-β on either proliferation or survival. Sensitivity to the TGF-β was heterogeneous between individuals and not significantly different between HSPCs

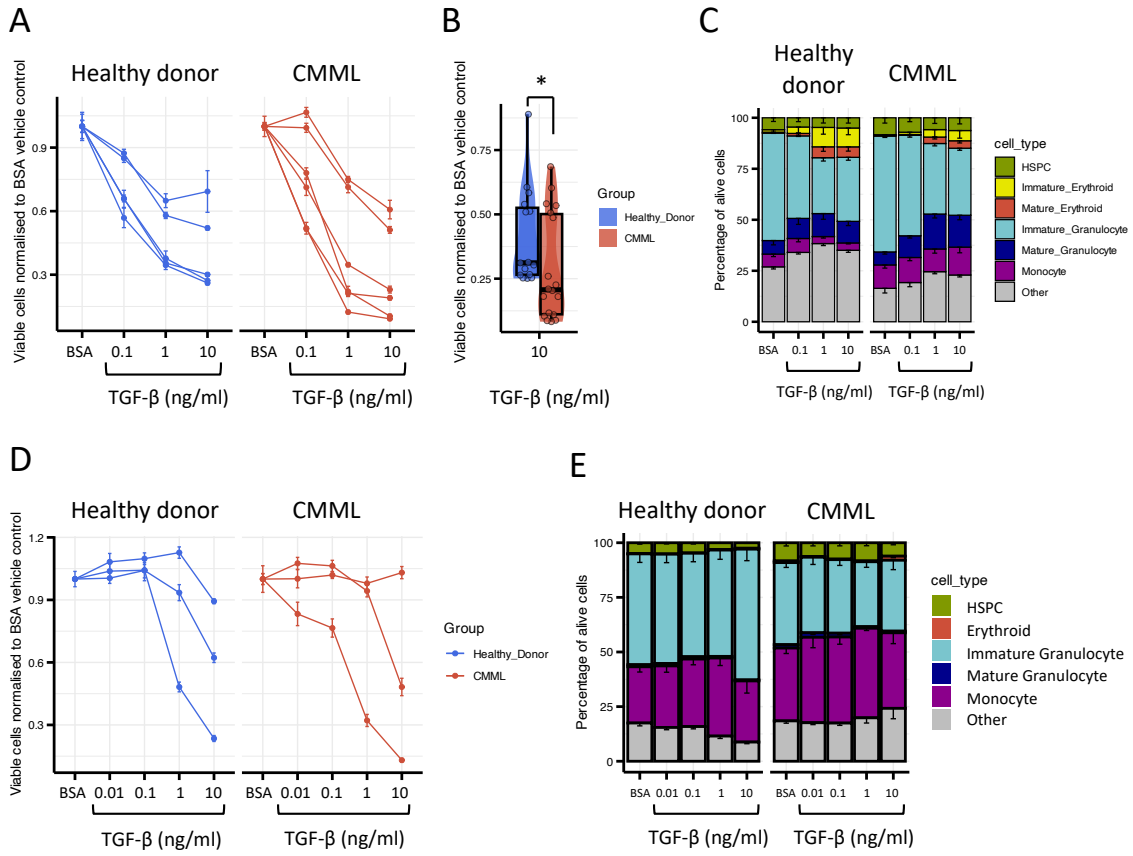


Figure 5.5: TGF-β causes dose dependent growth inhibition, suppression of myelomonocytic differentiation and enhanced erythropoiesis in TET2m CMML and healthy donor HSPCs - (A) Proliferation assay output. HSPCs were cultured for 8 days with a complete cocktail of proliferative cytokines and either BSA or 0.01 to 10ng/ml TGF-β. Viable cell counts were normalised to BSA cultures for each individual and split by disease group. (B) Box and whisker plot of 10ng/ml TGF-β supplemented cultures from (A). Each point represents an independent culture replicate. (C) Average percentage of each cell type by individual. (D) Survival assay output. HSPCs were cultured for 48 hours with low dose SCF, FLT3L and specified BSA/TGF-β concentration before addition of a complete prolfierative cocktail of cytokines and 11 further days of culture. Viable cells at D13 plotted by individual. Counts are normalised to BSA control cultures. (E) Average percentage of each cell type by individual for differet dose levels of TGF-β. Error bars represent standard error of means for all plots.

from CMML patients and healthy donor controls. This is therefore unlikely to be an important axis controlling HSPC dysfunction in TET2m CMML.

5.2.6 Emactuzumab inhibits proliferation and monocyte differentiation of HSPCs in M-CSF responsive TET2m CMML patients

The effect of M-CSF on HSPC proliferation was tested using a similar assay to that used in IL-1β, although instead of different concentrations of cytokine, different concentrations of emactuzumab were added to culture. In BSA vehicle treated cultures TET2m CMML HSPCs were more proliferative than healthy donors with mean cell counts of 218135 and 90328 respectively (p=0.02, Wilcoxon test). This difference was due to expansion in immature granulocytes (146359 vs 48110, p=0.009) and monocytes (16355 vs 594, p=0.0004). No difference was found between other cell populations, although there was a trend towards an increase in HSPC numbers (1113 vs 766, p=0.08). HSPCs in 4/7 (57%) TET2m CMML patients had an additional proliferative response when cultured with 100ng/ml M-CSF (CMML_responder). This was reversed in all patients by the addition of 100ng/ml of emactuzumab. Only 1/5 (20%) healthy donors had a proliferative response, classified as a greater than 20% increase in average cell number during 8 days of culture (Figure 5.6A and B). I compared TET2m CMML patients who were high M-CSF responders with TET2m CMML patients who were non-responders in BSA vehicle treated cultures and found that there was no difference in total alive cells, monocytes or immature granulocytes, however HSPCs were increased (1376 vs 753, p=0.02). In cultures treated with M-CSF, high M-CSF responders had a significant increase in monocytes (24592 vs 3784, p=$5x10^{-5}$) but this was not seen in patients who were non-responders (92229 vs 33640, p=0.65), although there was heterogeneity between individuals within the group. For instance, in one individual non-responder, 001/300, the total number of cells did not increase in response to treatment with M-CSF, but the proportion of monocytes being produced dramatically increased with a concomitant reduction in the proportion of immature granulocytes. Emactuzumab reversed this effect, but

did not result in a reduction of overall cell number in this patient. This indicates increased sensitivity to the effect of M-CSF on differentiation but not proliferation.

I compared demographic details of M-CSF responders and non-responders and found that all four responders had a pathological SRSF2 mutation but no non-responders (chi-square test p=0.06). There were no other differences detected between groups by chi-square or multivariable logistic regression analysis. The association with SRSF2 mutation should also be interpreted with caution given the small number of samples tested. Interestingly two of the responders were also found to have increased CSF1R expression on classical monocytes and both of these individuals were found to have an erythroid-biased HSPC phenotype by flow cytometry. Although no other individuals had CSF1R expression tested on classical monocytes, all of the samples cultured during the proliferation assay were tested for HSPC subpopulations during the sort. Re-analysis of these data found that M-CSF responsive patients had a higher percentage of MEPs as a total of lineage negative CD34 positive HSPCs compared to non-responsive patients, but the difference was not statistically significant (22.9% vs 10.6%, p=0.23). These data are suggestive of erythroid-biased CMML patients, as defined in Chapter 3, having increased sensitivity to M-CSF signalling. This is in keeping with the predictions of multinichenet and the hypothesis that the accumulation of M-CSF expressing erythroid progenitors in these patients was supporting the development of CMML. These assertions however remain speculative given the small number of patients used in this study and require further investigation to support or refute their validity.

HSPC survival was assessed after culture for 48 hours in serum free conditions with BSA or M-CSF and multiple concentrations of emactuzumab or vehicle control. In addition to the effect of M-CSF on HSPC survival, this assay also allowed the assessment of any inherent differences in HSPC survival in cytokine free culture between healthy donors and TET2m CMML. Healthy donor HSPCs were more resistant to cell death during cytokine free culture compared to TET2m CMML HSPCs, as demonstrated by an increased alive cell count at the end of the assay in BSA and emactuzumab vehicle treated cultures (324034 vs 239312,

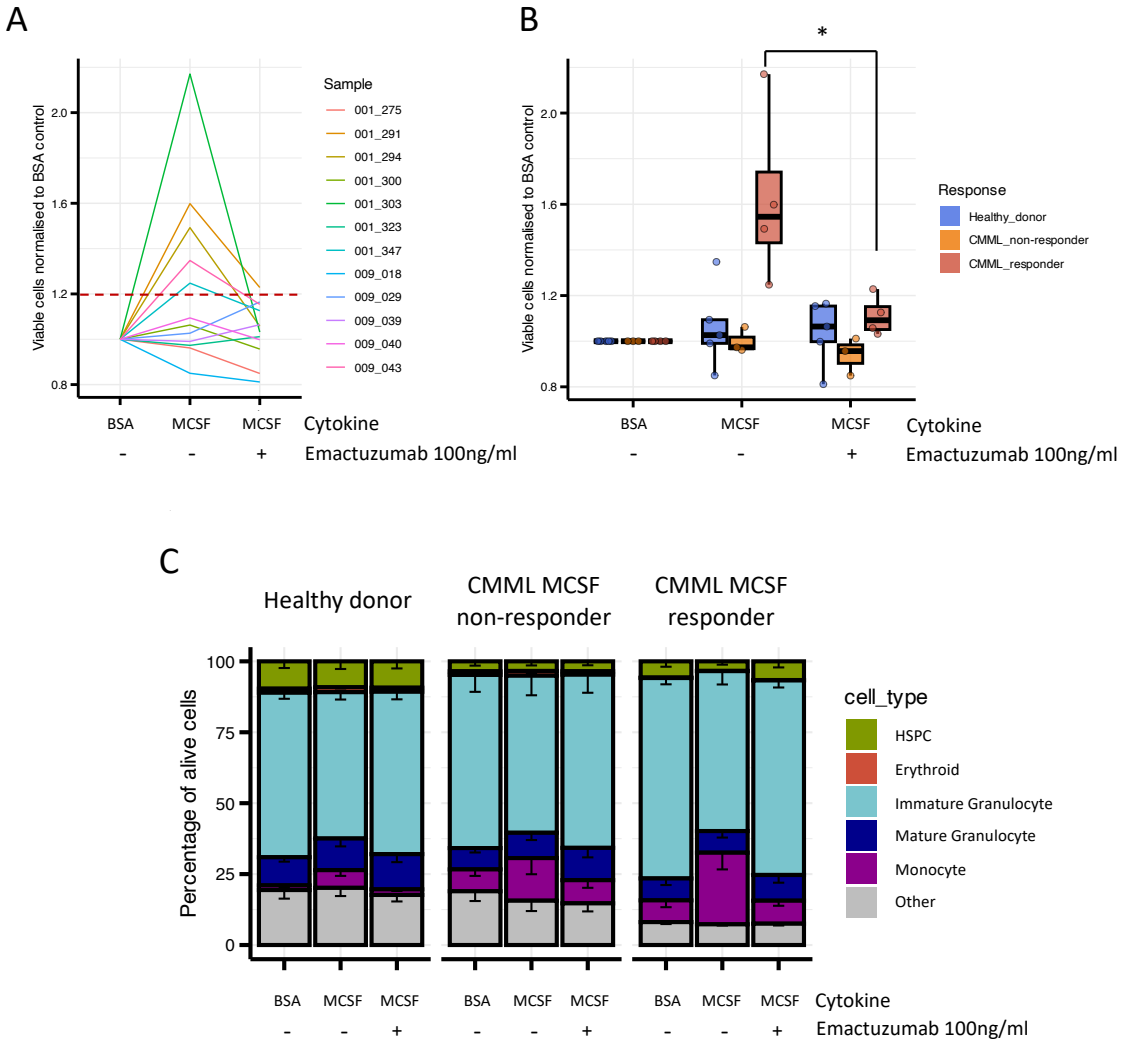


Figure 5.6: The majority of TET2m CMML patients exhibit HSPC hypersensitivity to M-CSF, which is completely reversed by emactuzumab - (A) Heterogeneous response to M-CSF stimulation is displayed by individual, with average cell count normalised to BSA vehicle wells. Individuals can be grouped into responders, which have >20% increase in cell number in M-CSF arm (above red dotted line) and non-responders (below red dotted line). (B) Summary of average cell count for each individual when grouped by response. (C) Cell type of detected cells, displayed as an average percentage of alive cells between individuals in each group. Error bars represent standard error of means for all plots.

p=0.02). There was a trend towards increased monocyte counts (48325 vs 64798, p=0.48) and decreased immature granulocyte counts (125600 vs 72705, p=0.059) in TET2m CMML patients, but no cell populations were significantly different. When comparing M-CSF to BSA supplemented cultures, healthy donors had a reduced alive cell count after culture with M-CSF (324034 vs 290083, p=0.008), which was reversed by the addition of 10μg/ml emactuzumab (290083 vs 343640, p=0.004). However, there was no such difference in cell count in M-CSF compared to BSA supplemented cultures in TET2m CMML patients (239312 vs 220182, p=0.64). However, this lack of difference is likely to be due to interpatient heterogeneity, with one patient demonstrating a 1.5-fold increase in M-CSF treated cultures and the other two patients following a similar pattern to healthy donors (Figure 5.7A). There was only one patient who had HSPCs tested in both the proliferation and survival assays: Patient 001/300. HSPCs had increased sensitivity to the effect of M-CSF on monocyte differentiation, however, M-CSF had a negative impact on HSPC survival, which could be reversed by emactuzumab.

M-CSF exerted a different effect on specific HSPC subpopulations, with M-CSF supporting monocyte-primed HSPC survival at the expense of granulocyte-primed HSPCs (Figure 5.7B). Both healthy donors and TET2m CMML had reduced mature granulocyte production in M-CSF supplemented cultures (10068 vs 3835, p=0.004, and 6942 vs 1527, p=0.008 respectively), and TET2m CMML patients had a trend towards an increase in monocyte count (64798 vs 104350, p=0.08). When comparing M-CSF vehicle supplemented HSPCs with M-CSF and 10μg/ml emactuzumab treated HSPCs, as shown in Figure 5.7C, TET2m CMML HSPC count was increased with emactuzumab (8148 vs 19391, p=0.02) and monocytes were decreased (104350 vs 61850, p=0.007). Healthy donors also demonstrated a decrease in monocyte count (59099 vs 30300, p=0.004) as well as an increase in immature granulocytes (158458 vs 224614, p=0.004).

In summary, M-CSF had a proliferative effect on 4/7 TET2m CMML patients, all of whom had additional SRSF2 mutations, and emactuzumab was able to reverse this effect at 100ng/ml. M-CSF supported the survival of monocyte-primed HSPCs

and reduced the survival of granulocytic-primed HSPCs in both healthy donors and TET2m CMML. Emactuzumab at 10μg/ml completely reversed this effect. The addition of M-CSF led to a reduction in total HSPC survival, with differences being most marked between TET2m CMML patients. This is possibly due to differences in proportions of monocyte-primed HSPC but more patients need to be tested to validate this assertion. Further work is now required to understand which patients are M-CSF responsive and to test emactuzumab in a model that is more representative of the bone marrow micro-environment, such as a murine xenograft or human organoid model.

5.3 Discussion

During the previous two chapters I have demonstrated that TET2m CMML HSPCs are primed for myelomonocytic differentiation by utilising the emergency myelopoiesis pathway controlled by the CEBPB/BACH2 molecular switch. This permissive epigenetic environment should result in hypersensitivity of HSPCs to external signals promoting myelomonocytic differentiation. I identified three candidate interactions and designed *in vitro* culture assays to investigate the impact on these interactions on HSPC differentiation in TET2m CMML. Additionally, in the case of CSF1R, the effect on monocyte proliferation was tested. I found that TET2m CMML monocytes expressed CSF1R and NCAM1, potentially making them targetable by immunotherapy. Blocking CSF1R signalling using emactuzumab produced a dose dependent reduction in monocyte proliferation but there was no difference in response between TET2m CMML patients and healthy donors. This finding was somewhat surprising given that CSF1R expression was increased on CMML monocytes. CSF1R expression is usually increased in atypical monocytes compared to classical monocytes and CSF1R inhibition has previously been shown to preferentially target atypical monocytes *in vivo*[431]. It is possible that the lack of differential response could be due to the culture conditions or culture duration. Furthermore, as demonstrated in the previous chapter, monocyte numbers are increased by 10-fold in TET2m CMML BM-MNCs and therefore a similar level of

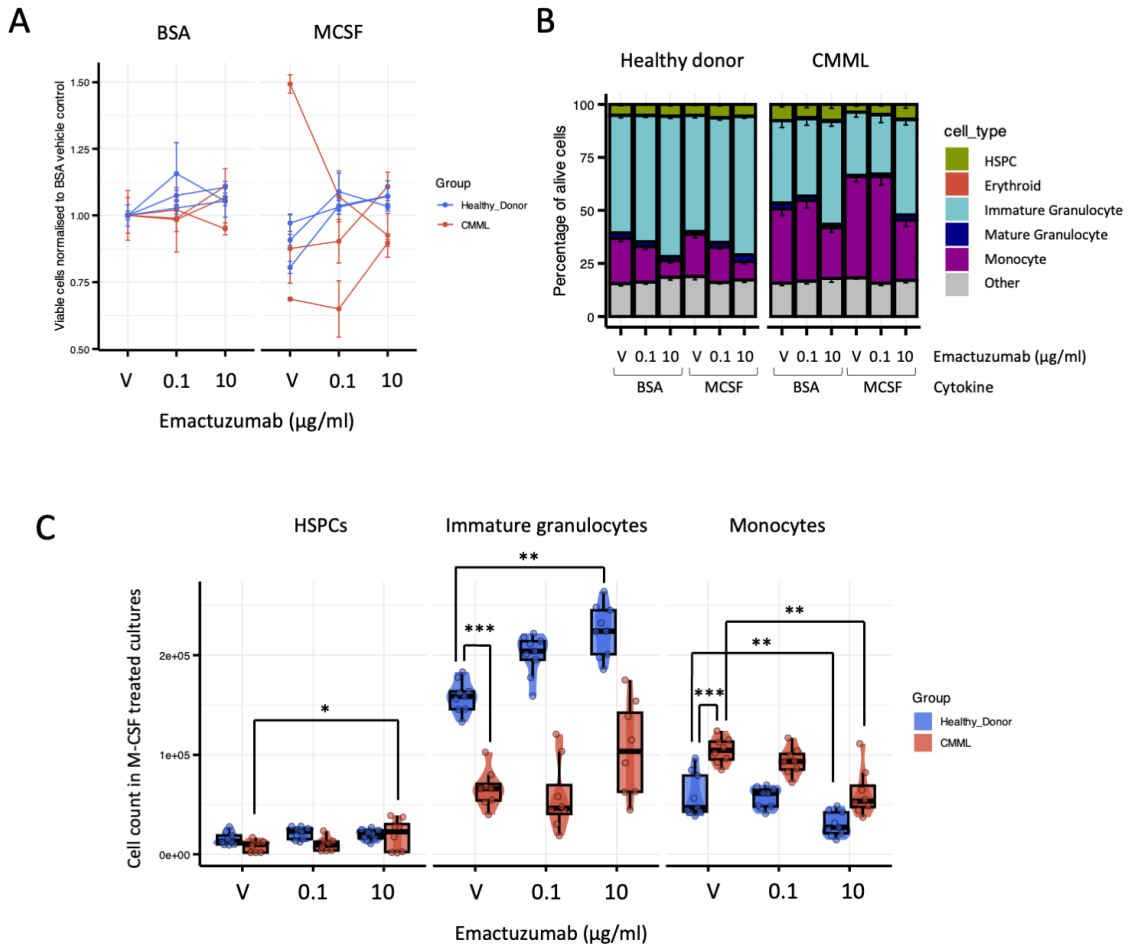


Figure 5.7: M-CSF promotes survival of monocyte-primed HSPC at the expense of granulocyte-primed HSPC - HSPCs were cultured for 48 hours in the presence of either M-CSF or BSA with multiple dose levels of emactuzumab. After 48 hours a proliferative cocktail of cytokines was added and the output of 11 further days of culture was assessed by flow cytometry. (A) Total number of viable cells at the end of culture is plotted by individual, normalised to BSA vehicle control wells and split by BSA and M-CSF treated cultures. (B) Average percentage of each cell type by individual, split by group to compare healthy donors to TET2m CMML. (C) Cell count output from M-CSF treated cultures demonstrating the different effect of emactuzumab on specific cell types. Error bars in A and B display standard error of means. Differences between groups were identified using Wilcoxon test. Significant ($p<0.05$) differences are displayed using brackets.

inhibition will have a far greater effect on monocyte numbers in CMML compared to healthy donors. Therefore, emactuzumab may become an exciting novel treatment through modifying monocyte numbers in CMML and thereby modifying the bone marrow micro-environment that is supporting HSPC differentiation.

57% of TET2m CMML patients tested had HSPCs that were hypersensitive to M-CSF, inducing a proliferative response. Another individual also demonstrated increased sensitivity, where M-CSF enhanced differentiation towards monocytes, rather than an effect on proliferation. Both the proliferative effect and the differentiation bias were reversed by emactuzumab, demonstrating that these effects are CSF1R specific and that this is a potentially targetable axis in CMML patients. M-CSF had a negative impact on HSPC survival in nearly all individuals when HSPCs were cultured in the absence of any other cytokine. However, a single patient demonstrated a survival advantage when M-CSF was added to the culture. Greater numbers of patients need to be tested to determine whether this finding was an anomaly or whether this may be an additional mechanism providing HSPC fitness advantage in certain individuals. The overall trend in HSPC survival is likely to reflect a variable effect of M-CSF on HSPC subtypes, with a negative impact on granulocyte-primed HSPCs and a supportive effect on monocyte-primed HSPCs. The design of this survival assay was focussed on isolating the M-CSF specific effect on HSPC survival, however this is not physiological and ignores the complex interplay between cytokines within the bone marrow micro-environment that help support HSPC survival and differentiation. M-CSF has been found to act synergistically with IL-3, GM-CSF and IL-6 in supporting HSPC colony formation[452], [453]. Therefore, analysis of the effect of CSF1R inhibition in a more physiological context such as a murine xenograft or a bone marrow organoid model is important before gauging the likely efficacy of this as a treatment of patients with CMML.

Further work is required to identify biomarkers for patients who are hypersensitive to M-CSF and would therefore potentially gain the greatest benefit from treatment with emactuzumab. All patients with a proliferative response to M-CSF had pathogenic mutations of SRSF2, a member of the spliceosome complex. SRSF2

mutations alter spliceosome mRNA binding motif specificity, resulting in recurrent mis-splicing of haematopoietic regulators such as EZH2[454]. It remains to be seen how SRSF2 mutation can increase the sensitivity of cells to CSF1R and this warrants further investigation. The only two patients included in both the HSPC proliferation assay and monocyte culture assay were both M-CSF responsive and had increased CSF1R expression on classical monoyctes, highlighting this as a potential biomarker. M-CSF responsive patients also had a trend towards over-representation of megakaryocyte-erythroid progenitors (MEPs) within the total HSPC population. This finding is interesting because M-CSF is produced by MEPs and early erythroid progenitors in healthy donors and CMML patients, which could create a feed-forward loop supporting HSPC proliferation.

TGF-β exhibited a dose dependent inhibitory effect on haematopoiesis in both the proliferation and survival assays with variability between individuals. TGF-β promotes HSC quiescence[379], and therefore may have been expected to promote increased survival in these cultures. However, HSCs require very specific culture conditions, particularly benefiting from hypoxia and semi-solid growth media[332]. Therefore, the lack of survival benefit in TGF-β cultures may in part be due to the normoxic and liquid culture conditions used in this assay. TGF-β is also known to have a negative impact on myelopoiesis[380], and the culture conditions used in both assays were most supportive of myeloid differentiation, which may also have contributed to these findings. However, the conditions were specifically designed to identify whether TET2m CMML patients were more resistant to this effect than healthy donors and this was not the case in either assay. The inhibition of HSPC proliferation was in fact greater in TET2m CMML patients compared to healthy donors at the highest dose level of TGF-β, 10ng/ml. This is at odds with the findings from single cell gene expression analysis and indicates that this is unlikely to be an axis that can be effectively targeted in CMML.

TGF-β also enhanced erythroid maturation in both healthy donors and TET2m CMML patients in a dose dependent manner. Previous studies suggest that TGF-β promotes terminal differentiation of erythroid progenitors, but has a negative

effect on proliferation of early erythroid progenitors[436]. The data presented here provides further evidence for this, at a dose of 5 or 10ng/ml TGF-β induced terminal differentiation of erythroid progenitors. However, it did not support erythroid progenitor survival. It should be noted that erythroid progenitor survival was in general low and therefore these conditions were not optimal to detect this. Ligands within the TGF-β family have been implicated in dyserythropoiesis and ligand traps are a promising therapeutic area for improving erythropoiesis in MDS and thalassaemia. However, their efficacy in CMML is currently unknown, as CMML patients were not included in registry studies for either luspatercept or sotatercept. There are also no current plans to find out, with no active studies registered on clinicaltrials.gov.

IL-1β had a dose dependent proliferative effect on both healthy donor and TET2m CMML HSPCs. The lack of difference between these groups indicates similar levels of sensitivity. However, differences in IL-1β concentration in the bone marrow micro-environment could still be a driver of TET2m CMML expansion. I was unable to determine whether this was the case in my IMC experiment as the antibody used for staining IL-1β had suboptimal performance. However, in a study of changes to cytokine concentrations in CMML patient serum, IL-1β was increased in patients with a driver mutation within an epigenetic modifier, such as TET2[109]. This is therefore an interesting axis to explore further, especially as many established treatments exist to inhibit this axis. Anakinra, an IL-1 receptor antagonist, and canakinumab, a monoclonal antibody that targets IL-1β, are frequently used in treatment of rheumatological disorders and there is increasing interest for using these agents in oncology. Preclinical evidence of their use in CMML exists, in *Tet2* knockout mice, additional knockout of *Il-1r1* rescues many of the haematological features[185]. Furthermore, NLRP3 inflammasome mediated IL-1/*beta* signalling is important in the pathogenesis of MDS, and a phase II study has demonstrated that this pathway can be effectively targeted by canakinumab[404], [455]. Therefore, although my findings are limited by small numbers and did not use a targeted

inhibitor, these results support further work in this area to test whether targeting this axis may be effective in CMML.

5.4 Conclusion

In vitro modelling of HSPC function in TET2m CMML successfully demonstrated hypersensitivity to M-CSF in 57% of TET2m CMML patients. This is in keeping with our hypothesis that permissive epigenetic changes support the C/EBP-β driven emergency myelopoiesis in TET2m CMML, priming HSPCs for proliferation and myelomonocytic differentiation in response to extrinsic cytokine signalling. Further work is required to identify whether this remains the case when HSPCs are cultured in a more physiological micro-environment and the development of a human bone marrow organoid system will be a major focus of my future work. This system will allow the testing of emactuzumab, as well as directly inhibiting C/EBP-β activity using lucicebtide, to provide preclinical evidence that targeting this axis may normalise CMML patient haematopoiesis and support the development of a phase II clinical trial.

6

Discussion

Contents

6.1 TET2 mutation creates a permissive epigenetic landscape for emergency myelopoiesis 188

6.2 Expanded mature myeloid populations create a feed forward loop to enhance emergency myelopoiesis . . . 191

6.3 Extrinsic cytokine signalling can be targeted to correct HSPC dysfunction in TET2m CMML 192

6.4 Implications and wider relevance 194

6.5 Limitations 195

6.5.1 Is deep molecular phenotyping of patient samples the correct approach? 195

6.5.2 Limitations of chosen techniques 197

6.6 Future Work 198

6.6.1 How does TET2 dysfunction impact cell metabolism and is this a targetable vulnerability? 198

6.6.2 What is the mechanism through which TET2 loss of function promotes C/EBP-β activity? 199

6.6.3 Can the aberrant epigenetic priming be targeted to mitigate HSPC dysfunction in CMML? 199

6.6.4 How does TET2 loss of function contribute to the pathogenicity of co-mutations? 200

6.7 Concluding remarks 201

6.1 TET2 mutation creates a permissive epigenetic landscape for emergency myelopoiesis

CMML is characterised by the proliferation of myeloid-biased progenitors causing excess production of myelomonocytic immune cells. TET2 mutation is the most common founder mutation, but little is known about how TET2 loss of function can contribute to the pathogenesis of CMML. I present data demonstrating that TET2m CMML patients can fit into either of the erythroid-biased, myeloid-biased or balanced HSPC groups that were described using scRNA-seq by *Ferral-Fairbanks et al*[17]. I found that bulk chromatin accessibility was able to distinguish these groups but not bulk gene expression analysis and that these differences stemmed from a change to the make up of HSPC subpopulations found in patients belonging to these different groups.

Despite these differences, gene expression analysis found that HSPCs in all groups were enriched for expression of genes involved in oxidative phosphorylation and proliferation, indicating that TET2m CMML HSPCs share metabolic activation regardless of HSPC differentiation bias. This is potentially an important finding, as increased reliance on oxidative phosphorylation may be targeted by specific pathway inhibitors. Dependence on oxidative phosphorylation has recently been found in HSPCs from patients with DNMT3A^{R882H} mutations, which like TET2 are recurrently found in clonal haematopoiesis and myeloid malignancies. This metabolic dependency can be exploited using metformin, a commonly used anti-diabetic medication[456]. In this study they used population screening data to correlate clonal haematopoiesis incidence in diabetic patients on metformin and found that DNMT3A^{R882H} driven clonal haematopoiesis was reduced. However, clonal haematopoiesis secondary to mutations in other genes, including TET2, were not altered. This study focussed on pre-malignant clonal haematopoiesis, whereas in TET2m CMML, where HSPC proliferation is likely to further increase reliance on mitochondrial metabolism, metformin may have greater efficacy. Therefore, investigation of repurposing metformin for use in TET2m CMML should be strongly considered.

To gain further understanding of the mechanisms underpinning the expansion of specific HSPC subgroups in TET2m CMML patients I conducted single cell chromatin accessibility and gene expression analysis. This found evidence that lineage priming was initiated by changes within HSCs, the earliest stage of differentiation. However, HSCs from both erythroid-biased and myeloid-biased patients shared increased chromatin accessibility of C/EBP family motifs. This indicates that the activity of C/EBP family TFs are likely to be important in the pathogenesis of the disease irrespective of a patient's HSPC phenotypic group. C/EBP-α is one of the most important regulators of steady state myelopoiesis and loss of function mutations in this gene are recurrently found in AML, causing differentiation arrest[457]. C/EBP-β also controls myeloid differentiation, however its function is redundant in steady state haematopoiesis, but is fundamental in emergency myelopoiesis in response to inflammatory insult. Detection of an inflammatory challenge, such as exposure to lipopolysaccharide, causes chromatin accessibility to be altered within HSCs, forming a heritable epigenetic innate immune memory in these cells. This facilitates more rapid engagement of C/EBP-β driven emergency myelopoiesis in response to subsequent stimulation[298]. C/EBP-β was found to be particularly active in both erythroid-biased and myeloid-biased TET2m CMML on footprint analysis of HSCs and pseudotime analysis found that both *CEBPB* gene expression and chromatin accessibility of C/EBP-β targets were increased in HSPCs throughout early differentiation. These findings are indicative of innate immune memory forming in TET2m CMML HSPCs and indicate that HSPCs are primed for myelomonocytic differentiation in response to extrinsic cytokine signalling.

C/EBP-β forms a molecular switch with BACH2, a negative regulator of myelopoiesis and promoter of erythropoiesis. I used figR to correlate gene expression to chromatin accessibility and infer TF activity by motif enrichment analysis. This created a comprehensive overview of gene regulatory networks controlling haematopoiesis in TET2m CMML. In this analysis BACH2 activity was reduced in both erythroid-biased and myeloid-biased TET2m CMML patients. C/EBP-β and BACH2 bind mutually exclusively to genomic loci and exhibit different

avidity to 5-mC containing motifs. C/EBP-β binding avidity is increased by the presence of 5-mC and BACH2 decreased[146]. This implicates, for the first time, TET2 as a regulator of emergency myelopoiesis. I hypothesise that TET2 actively removes methylation from C/EBP-β target loci and thereby switches off emergency myelopoiesis. This has broad implications for the study of cancer, where TET2 loss of function mutations are commonly found, as well as other areas such as critical illness, where DNA hypermethylation has been linked with monocyte exhaustion and resultant immune dysregulation[458].

Erythroid-biased TET2m CMML patients had disruption to gene regulatory networks controlling terminal erythropoiesis and were enriched for patients with SRSF2 co-mutation. Gene expression during terminal erythropoiesis is regulated by a dynamic process of alternative splicing[459]. SRSF2 mutations, such as the P95H mutation, alter the RNA binding avidity and motif recognition of SRSF2, resulting in aberrant exon inclusion events during splicing[162], [454]. SRSF2 mutations disrupt the normal functioning of terminal erythropoiesis, resulting in macrocytic anaemia[454]. However, SRSF2^{P95H} mice do not develop a CMML-like phenotype unless *Tet2* is additionally knocked out, indicating that an interaction between the co-mutations is required for myelomonocytic expansion[460]. This raises the question, how do SRSF2 and TET2 mutations interact? They may have different effects, such as SRSF2 mutation causing aberrant splicing of critical transcripts for promoting erythroid differentiation, which in turn allows TET2 dysfunction to promote myelomonocytic differentiation of multipotent progenitors. Alternatively they may have interdependent effects, such as SRSF2 mutation altering the binding avidity to methylated m5C, a target for TET2 oxidation, on pre-mRNA molecules. Further work is required to elucidate the exact mechanisms underlying this interaction.

6.2 Expanded mature myeloid populations create a feed forward loop to enhance emergency myelopoiesis

Having established that HSPCs are epigenetically primed for emergency myelopoiesis, I sought to identify potential cytokine triggers and cell types that were responsible for producing this signal. HSPCs reside within highly specialised niches that provide supportive cytokine signals that are specific for their needs. In TET2m CMML I found that the normal cell types within this niche were disrupted, with increased monocyte and macrophage populations and decreased HSPCs, mature granulocytes, T cells, erythroid cells and endothelial cells. The change in composition of cells within the niche is highly likely to influence the cytokine signals being received by HSPCs, which can alter their function. I conducted a multiomic analysis of bulk sorted monocytes and found that TET2m CMML monocytes were more proliferative, metabolically active and were less pro-inflammatory. Published data on CMML monocytes is conflicting, with one study finding similar changes to those presented here[366], and another, where they sought to identify CMML specific changes by controlling for age, finding CMML monocytes to be pro-inflammatory[110]. Using single cell multiomic analysis I was able to explain these conflicting findings. CMML patients have expansion of immature monocytes and monocytic progenitors, which are inherently more proliferative and produce less pro-inflammatory cytokines. Therefore bulk sorting samples on CD14 expression will capture these immature monocytes as well as mature monocytes, leading to an overall immature picture in the bulk analysis when compared to healthy donors. However, when comparing single cell transcriptomically annotated mature monocytes between TET2m CMML and healthy donors, patient monocytes were pro-inflammatory, with enrichment for interferon response and inflammatory signalling pathways. Therefore the differences in monocyte enrichment strategy or possibly the removal of age related differences from the analysis allowed for a cleaner analysis of mature monocytes in the study by *Franzini et al*[110].

TET2m CMML monocytes had increased expression of HGF, a growth factor implicated in AML pathogenesis[369], and multiple adhesion molecules such as DSC2, NCAM1 and NRXN1, which may account for their increased presence in the HSPC niche. Cell-cell communication analysis using single cell gene expression identified several critical ligand-receptor interactions with HSPCs that were predicted to promote the HSPC dysfunction seen in TET2m CMML, such as CSF1-CSF1R, TGFB-TGFBR1 and IL1B-IL1RAP. Imaging mass cytometry confirmed the expression of CSF1/CSF1R and TGFB/TGFBR1, but was unable to confirm NCAM1, HGF or IL1B due to poor antibody staining or lack of expression. CSF1R and NCAM1 expression were however increased on *ex vivo* cultured TET2m CMML monocytes when assessed by flow cytometry. I was therefore able to identify the presence of ligand and receptor protein expression for multiple potential interactions between monocytes and HSPCs in TET2m CMML. I also found that monocyte populations were expanded by 10 times in TET2m CMML BM-MNCs, making monocyte-HSPC interactions of particular interest for promoting HSPC dysfunction. These interactions may also help create a feed-forward loop by stimulating emergency myelopoiesis in epigenetically primed TET2m CMML HSPCs, resulting in further monocyte production.

6.3 Extrinsic cytokine signalling can be targeted to correct HSPC dysfunction in TET2m CMML

Cytokines are essential for supporting survival, promoting proliferation and coordinating HSPC differentiation to meet the specific haematopoietic requirements of an organism. This inherent adaptability of HSPCs to extrinsic signalling is essential for reacting to insults such as infection or bleeding. However, this same adaptability can be exploited in carcinogenesis to provide a malignant clone with a competitive advantage. I tested the effects of TGF-β, IL-1β and M-CSF on TET2m CMML HSPCs to identify whether they provide an advantage either by increasing proliferation, enhancing survival or promoting monocytic differentiation. This found that TGF-β had a dose dependent negative effect on HSPC proliferation and

survival for both healthy donor and TET2m CMML HSPCs. Whereas, IL-1β had a proliferative effect on both healthy donor and TET2m CMML HSPCs. Although there was no difference in the proliferative response between the groups, this is potentially a vulnerability that may be exploited, because IL-1β has increased expression in TET2m CMML[109]. Assessment of IL-1β expression by IMC in this study was inconclusive due to the poor performance of the antibody, however, TET2m CMML patient serum contains elevated IL-β and evidence is emerging that this may be an effective axis to target in CMML and MDS[404], [461]. Furthermore, TET2m CMML HSPCs displayed increased gene expression of *IL1RAP*, which has been confirmed using flow cytometry, making IL1RAP an exciting potential target of immunotherapy[360]. Cell surface expression of IL1RAP has also been found on AML and CML leukaemic stem cells, and can be successfully targeted using BOS-371, a humanised monoclonal antibody, and BIF002, an anti-IL1RAP/CD3 bispecific T cell engager, in pre-clinical studies[462]–[465].

57% of TET2m CMML patient HSPCs had a hypersensitive proliferative response to M-CSF, with additional patients having enhanced monocytic differentiation and improved survival. These effects were completely reversed by the addition of emactuzumab, confirming they are CSF1R specific. M-CSF is present at increased levels in CMML patient serum and I found increased CSF1R expression on TET2m CMML patient monocytes[109]. Furthermore, M-CSF, along with IL-4, are well established drivers of macrophage polarity towards an M2 immunosuppressive phenotype. Recent analysis of CMML monocyte markers of polarity found them to be M2 in phenotype, suggesting that M-CSF may be important in dictating their phenotype[366]. These data indicate the emactuzumab may be an exciting treatment that could reduce CMML HSPC fitness advantage, reverse myelomonocytic differentiation bias and correct monocyte/macrophage polarity that is contributing to immune dysregulation in CMML. These would each have meaningful impact on patient quality of life and perhaps reduce disease burden by supporting wild type haematopoiesis recovery. However further work is needed to test for the

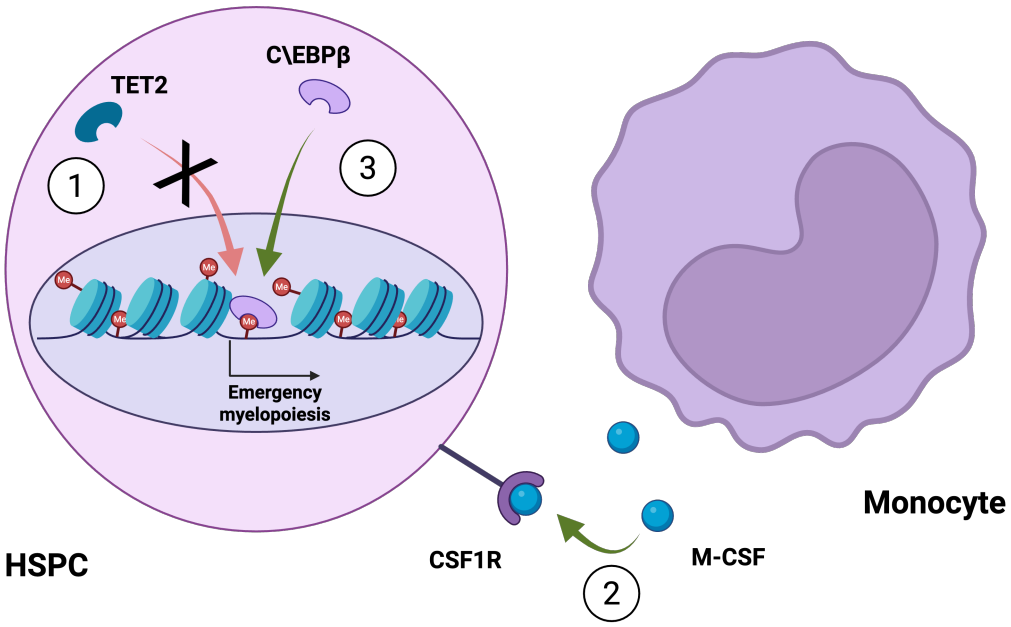


Figure 6.1: Visual Summary - TET2 dysfunction epigenetically primes HSPCs for myelomonocytic differentiation using an emergency myelopoiesis-like response to extrinsic cytokine signalling - (1) TET2 dysfunction leads to epigenetic priming of HSPCs. (2) Extrinsic signals, such as M-CSF, trigger (3) C/EBP-β activity, promoting myelomonocytic differentiation through the emergency myelopoiesis pathway.

effect of CSF1R inhibition in a physiologically meaningful environment such as a xenograft or bone marrow organoid model.

6.4 Implications and wider relevance

For the first time I have demonstrated that TET2 mutant CMML HSPCs provide a supportive epigenetic environment for the activity of C/EBP-β, which also suppresses the activity of its antagonist, BACH2. This promotes myelomonocytic differentiation of HSPCs, utilising an emergency myelopoiesis pathway. These findings open up a raft of new potential therapeutic targets to be investigated in CMML, either by inhibiting C/EBP-β activity or extrinsic signals that are known to trigger EM. Further work is required to confirm whether this is secondary to altered methylation of binding motifs, but if this were to be found it would

link TET2 activity with the regulation of inflammation through a completely new mechanism, which would have broad reaching implications to the study of infection and autoimmune diseases as well as cancer.

6.5 Limitations

6.5.1 Is deep molecular phenotyping of patient samples the correct approach?

Deep molecular phenotyping of patient samples remains a controversial approach to scientific discovery. In many ways it is at odds with hypothesis driven science, where a given theory is tested for falsehood either by deductive reasoning or experimentation. This has been the prevailing scientific method of the last two millennia since it was devised by Aristotle in the 1st century BCE and refined by Francis Bacon in the 17th century and Karl Popper in the 20th century[466]. However, we are living through a time of both technological and philosophical revolution in science, where the capacity to create and analyse data is completely unprecedented. There are particular parallels between next generation sequencing (NGS), the paradigm shifting technology of this century, and optics, the revolutionary technology of Bacon and Galileo's 17th century. The discovery of the telescope and the microscope allowed the natural world to be observed in unprecedented resolution and prompted a shift in the scientific method from the classical dogma of deductive reasoning to that of empirical observation and mathematical analysis. Similarly, the subcellular discoveries of nucleic acid sequences afforded by NGS, is prompting further refinement in the scientific method to accommodate such a ground breaking technology.

However, NGS is not the only technology that is likely to have such profound consequences this decade. We sit on the precipice of an era of artificial intelligence (AI) that is likely to have far broader and unpredictable consequences on the scientific method than NGS. These technologies are complementary and for many years repositories of NGS data have been developed to aid human scientists partaking in hypothesis driven research to interrogate the data for their own purposes. When

AI is able to start asking questions of such data it will not follow this approach for long. AI is already used in a limited capacity to help model data. In this thesis for instance I have used machine learning algorithms for specific tasks, such as to remove noise from my data created by batch effect or ambient RNA contamination. However, tools exist now that can facilitate entirely autonomous AI driven scientific discovery[467]. These tools are currently limited by the accuracy and novelty of their discoveries, but provide proof of principle that autonomous AI, using established large language models, can conduct scientific research independently of humans. This will require a paradigm shift in how we approach scientific discovery.

AI has the potential to tap into data repositories on an unprecedented scale and therefore the generation of carefully annotated multiomic data has far greater potential to aid future discoveries than the established scientific method using supervised human analysis. This transformation in how science is performed will require a similar transformation in how data is validated and published. Moves towards automation of drug screening platforms are likely to be required to keep pace with the targets generated by AI. We may also find that the current gold standard peer review process will require updating to improve the capacity to publish findings quickly. Furthermore, there will be challenges to the legal framework protecting intellectual property, as it is not clear whether this should be held by the individuals creating the original data used to train the models, the architects creating the AI algorithm or the scientists guiding the AI towards a question of their interest.

The coming decades promise to be an exciting time to be a scientist, full of challenges and opportunities. However, these opportunities will not be shared equally, with the power to use such powerful technologies being limited to those with the deepest pockets. We will probably continue to see the gulf between the scientific super powers of the West and the rest of the world deepen. If the discussions surrounding intellectual property rights, eluded to above, comes down in favour of the AI makers, then it is not inconceivable that medical research, where there are particular commercial interests, shifts from being predominantly performed in Universities to the offices of Google or Open AI. There are certainly signs of interest

in moving into this field from Google, who recently announced an expansion of their DeepMind programme, that currently runs the AlphaFold project solving protein structures, into gene regulation by developing a new tool called AlphaGenome.

6.5.2 Limitations of chosen techniques

Working with primary patient samples, and particularly samples from a rare cancer like CMML presents certain challenges. Due to the rarity of the disease, there is a limited pool of patients to recruit for sample collection. Furthermore, not every CMML patient has a bone marrow biopsy performed because of the frailty of the patient population and the absence of effective treatment options after a diagnosis is secured. This reduced the power to detect, and the statistical confidence, in the results from experiments. However, the numbers of patients I have included in my experiments are similar to those within the published literature, where only a single study has published gene expression data from a cohort of several hundred patients[112].

Imaging mass cytometry is a well validated technique for multiplexed imaging of histology samples. Heavy metal conjugated antibodies are used to stain samples that are then ablated using a $1\mu m^2$ laser. Vaporised metals are sucked into a mass cytometer to identify their mass and quantity for each pixel. The technique therefore has a maximum resolution of $1\mu m^2$, which is adequate for most tissues, but probably inadequate for bone marrow trephines where there is a high density of small cells. It is perhaps for this reason that very few imaging mass cytometry studies have been published using bone marrow trephine specimens. This lack of resolution resulted in challenges during cell segmentation and annotation and limited the ability for downstream unsupervised analysis. Despite this, visual assessment of unidentified cells found that they were uniformly distributed throughout the regions of interest, and therefore it is unlikely that this introduced systematic bias to this analysis.

In vitro liquid culture assays were designed to investigate the impact of ligand interactions and targeted inhibition on cell growth and differentiation. Liquid culture assays are convenient for the ease of set up, scalability and ability to read

out by flow cytometry or plate reader. However cells do not receive the supporting interactions with other haematopoietic and stromal cell populations normally found within the niche. This is likely to alter cell function and their response to specific stimuli or inhibition. Further work is therefore required using a physiologically relevant model, such as a human bone marrow organoid or xenograft mouse model, to provide sufficient pre-clinical evidence to support the development of a phase II clinical trial of emactuzumab in CMML.

6.6 Future Work

6.6.1 How does TET2 dysfunction impact cell metabolism and is this a targetable vulnerability?

Throughout my analysis of bulk HSPCs and monocytes and the analysis of single cell gene expression, genes involved in oxidative phosphorylation, glycolysis and fatty acid metabolism were consistently found to be upregulated in TET2m CMML patients. Confirmation of increased respiratory capacity of CMML monocytes has also recently been reported, confirming their increased utilisation of oxidative phosphorylation and validating the findings of gene expression analysis[366]. It is not yet clear whether this is a TET2m specific effect or whether it is a hallmark of CMML, however reliance on oxidative phosphorylation is a potentially targetable vulnerability and therefore urgent work is needed to characterise this further. The first step of this investigation would be to review the leading edge analysis of my data to pull out the specific genes that are responsible for enrichment of these metabolic pathways. Next would be to test specific inhibitors of these mitochondrial genes on cultured bone marrow mononuclear cells and measure which cell populations are dependent on oxidative phosphorylation. Work on TET2 knockout cell lines and TET2 mutant AML cells has identified a distinct, shared metabolic signature for TET2 loss of function[468]. However further work is needed to identify whether this is associated with oxidative phopshorylation and whether this signature identifies TET2m specific vulnerabilities.

6.6.2 What is the mechanism through which TET2 loss of function promotes C/EBP-β activity?

TET2 loss of function leads to hypermethylation of many loci of the genome and as outlined above this has the potential to alter the activity of C/EBP-β and BACH2. Furthermore, analysis of TET2m in clonal haematopoiesis found that hypermethylated regions were particularly found within areas associated with genes responsible for controlling myeloid immune cell differentiation and activation[180]. Therefore, analysis of alterations to methylation within these areas in TET2m CMML HSPCs would provide vital evidence linking the loss of TET2 activity with the epigenetic priming of HSPCs. I have gDNA and Tn5 treated DNA stored from TET2m CMML HSPCs and am planning to use Tet-assisted pyrimadine borane sequencing (TAPS-β) to map changes to methylation in these cells. To optimise the conditions for this experiment I have created TET2 knockdown THP1 and HEL cell lines and am currently working with Skirmantas Kriaucionis to perform preliminary experiments. Following on from this analysis I intend to track the specific changes to C/EBP-β activity and BACH2 activity by performing CUT&RUN sequencing on isogenic TET2 knockout cell lines. I will then use nanopore sequencing of the TF bound DNA to confirm altered methylation of these loci.

6.6.3 Can the aberrant epigenetic priming be targeted to mitigate HSPC dysfunction in CMML?

As outlined in the limitations section above, the 2D culture assays I employed for testing altered activity of TET2m CMML HSPCs was a biased analysis. The culture conditions were specifically created to isolate the effects of the added cytokine, but these assays were quite distinct from the normal physiological environment found within the bone marrow niche. Therefore, this is not a relevant assay in which to perform pre-clinical drug testing to support the development of a targeted inhibitor into human clinical trials. I have been in discussion with Dr Abdullah Khan, who has developed a human bone marrow organoid system, to collaborate in the creation of a pre-clinical drug screening model of CMML[469]. This system uses

human induced pluripotent stem cells that are pushed towards differentiation into the multiple lineages found within the bone marrow micro-environment, leading to the creation of vascular structures, supporting niche cells such as fibroblasts and mesenchymal stromal cells and haematopoietic cells. This organoid can then be seeded with CellTrace labelled patient HSPCs, which can then be sustained in culture for several weeks. This model has been validated to support both lymphoid and myeloid malignancies, including those that are traditionally difficult to sustain in culture. After establishment of the seeded organoid I will add emactuzumab or lucicebtide, a targeted small molecule inhibitor of C/EBP-β, and monitor the impact of these inhibitors on cell growth, differentiation and bone marrow organisation by flow cytometry and immunohistochemistry.

6.6.4 How does TET2 loss of function contribute to the pathogenicity of co-mutations?

Mutations in a small number of genes are recurrently found in clonal haematopoiesis and myeloid malignancies[25]. Despite this relatively limited genetic diversity, combinations of these mutations are able to produce diverse clinical phenotypes. Understanding how these mutations interact to create certain phenotypes is therefore of great interest to the field. Combinations of mutated genes do not occur at random, with particular co-mutations being associated with specific diseases. For instance co-mutation of TET2 and SRSF2 are enriched within CMML, although they are also found in AML and to a lesser extent in MDS and MPN[470]. Furthermore, mutations in spliceosome members are enriched in TET2m MDS, potentially indicating a specific role for TET2 in coordinating splicing events[471]. This has never previously been demonstrated and would have broad reaching implications.

I found that patients with SRSF2 mutations were enriched within TET2m CMML patients with an erythroid-biased HSPC phenotype and that this was associated with disruption of gene regulatory networks controlling terminal erythropoiesis. I will therefore re-analyse the bulk HSPC RNA-seq data to identify alternative splicing events and alternative transcript utilisation. SRSF2 mutations are particularly

associated with exon inclusion events and therefore this may manifest as an alternative transcript use rather than an aberrant splicing event[162]. I will interrogate this data looking for evidence of aberrant splicing or alternative transcript expression in genes associated with terminal erythropoiesis. TET2 has been recently identified as being important for oxidation of m5C on RNA molecules and disruption of this activity was found to contribute to leukaemogenesis[472]. Therefore, methylation may alter mutant SRSF2 binding leading to altered splicing events. To investigate this I will create a TET2 knockout $SRSF2^{P95H}$ isogenic cell line using erythroleukaemic K562 cells and perform native RNA-seq using nanopore sequencing to identify changes to splicing or transcript use that overlap with events found in the patient data. Nanopore sequencing is able to detect m5C during native RNA molecule sequencing and therefore I can analyse m5C percentage and position in these aberrant transcripts to determine whether an interaction between m5C and mutant SRSF2 is likely to be causative of these events.

6.7 Concluding remarks

In this thesis I used multiomic profiling of TET2m CMML patient samples to discover that HSPCs were epigenetically primed for myelomonoytic differentiation using the C/EBP-β dependent emergency myelopoiesis pathway. This rendered HSPCs hypersensitive to M-CSF, which is expressed by monocytes, macrophages and mesenchymal stem cells that reside in close proximity to HSPCs in the CMML bone marrow niche. I demonstrated that CSF1R signalling could be inhibited using emactuzumab, a licensed monoclonal antibody, abrogating the M-CSF induced proliferative response and reversing the monocytic bias of HSPC differentiation. Further work is needed, but these data provide proof of principal that emactuzumab could be used as a potential novel treatment for this orphan malignancy. I intend to build upon this work during my post-doctoral research with the ultimate goal of bringing emactuzumab into a clinical trial for CMML patients.

Appendices

A
Appendix

Contents

A.1 Supplementary Tables 205
A.1.1 SmartSeq3 reagents and volumes 205
A.1.2 10X multiome reagents and volumes 205

A.1 Supplementary Tables

A.1.1 SmartSeq3 reagents and volumes

A.1.2 10X multiome reagents and volumes

Table A.1: SmartSeq3 lysis buffer

Reagent	μl per reaction	Supplier
Polyethylene Glycol 8000 (40% solution)	1	Sigma-Aldrich (P1458)
Triton X-100 (10% solution)	0.06	Sigma-Aldrich (93443)
RNase Inhibitor (40 U/µL)	0.08	TakaraBio (2313B)
dNTPs (10 mM)	0.4	Thermo Scientific (R0192)
OligodT30VN (100 µM)	0.04	IDT (5'-biotin-ACGAGCATCAGCAG-CATACGAT30VN-3')
RT-PCR grade H20	0.42	Invitrogen (AM9935)
RNA	4	

Table A.2: SmartSeq3 reverse transcription mix

Reagent	μl per reaction	Supplier
Tris-HCl pH 8.3 (1M)	0.2	Invitrogen (15568025)
NaCl (1M)	0.24	Invitrogen (AM9760G)
MgCl2 (100mM)	0.2	Invitrogen (AM9530G)
GTP (100mM)	0.08	Invitrogen (18332015)
DTT (100mM)	0.2	Thermo Scientific (707265ML)
RNase Inhibitor (40u/uL)	0.1	TakaraBio (2314B)
Template switch oligo (100uM)	0.16	IDT (5'- biotin-AGAGACAGATTGCGCAAT-GNNNNNNNNrGrGrG)
Maxima H-minus RT enzyme (200U/uL)	0.08	Thermo Scientific (EP0751)
Nuclease Free Water	0.3	Invitrogen (AM9935)

Table A.3: SmartSeq3 PCR mix

Reagent	μl per reaction	Supplier
Kapa HiFi Fidelity buffer (5X)	4	Roche (KK2502)
dNTPs (10mM/each)	1.2	Thermo Scientific (R0192)
MgCl2 (100mM)	0.1	Invitrogen (AM9530G)
Fwd Primer (10uM)	1	IDT (5'-TCGTCGGCAGCGTCAGATGTGT-ATAAGAGACAGATTGCGCAA*T*G-3')
Rev Primer (10uM)	0.2	IDT (5'-ACGAGCATCAGCAGCATAC*G*A-3')
Polymerase (1U/uL)	0.4	Roche (KK2502)
Nuclease Free Water	5.1	Invitrogen (AM9935)

Table A.4: ATAC transposition buffer

Reagent	μl per reaction	Supplier
0.5% digitonin	0.25	Promega (G9441)
10% Tween-20	0.25	Bio-Rad (1662404)
RNase-free PBS	8.25	Invitrogen (AM9625)
Ultrapure Water	2.5	Invitrogen (10977049)
2x Tagment DNA buffer	12.5	Illumina (15027866)
TDE1 Tn5 transposase (after sort)	1.25	Illumina (20034198)

Table A.5: ATAC index PCR mix

Reagent	μl per reaction	Supplier
UltraPure Water	10	Invitrogen (10977049)
25μM Forward Custom Nextera Primer (Ad1noMX)	2.5	Biomers.net
25μM Reverse Custom Nextera Primer (Ad2.x)	2.5	Biomers.net
NEBNext High Fidelity 2x PCR Master mix	25	NEB (M0541L)
Transposed DNA	10	

Table A.6: 10X multiome lysis buffer

Reagent	μl per reaction	Supplier
1M Tris-HCl (pH 7.4)	5	Sigma-Aldrich (T2194)
5M NaCl	1	Sigma-Aldrich (59222C)
1M $MgCl_2$	1.5	Sigma-Aldrich (M1028)
10% BSA	50	Miltenyi Biotec (130-091-376)
10% Tween-20	5	Bio-Rad (1662404)
500mM DTT	1	Sigma-Aldrich (646563)
RNase inhibitor	25	Sigma-Aldrich (3335402001)
10% Nonidet P40 Substitute	5	Sigma-Aldrich (74385)
2% Digitonin	2.5	Promega (G9441)
UltraPure Water	411.5	Invitrogen (10977049)

Table A.7: 10X multiome wash buffer

Reagent	μl per reaction	Supplier
1M Tris-HCl (pH 7.4)	5	Sigma-Aldrich (T2194)
5M NaCl	1	Sigma-Aldrich (59222C)
1M $MgCl_2$	1.5	Sigma-Aldrich (M1028)
10% BSA	50	Miltenyi Biotec (130-091-376)
10% Tween-20	5	Bio-Rad (1662404)
500mM DTT	1	Sigma-Aldrich (646563)
RNase inhibitor	25	Sigma-Aldrich (3335402001)
UltraPure Water	411.5	Invitrogen (10977049)

Table A.8: 10X multiome diluted nuclei buffer

Reagent	μl per reaction	Supplier
Nuclei Buffer (20X)	50	10X Genomics (2000153)
500mM DTT	2	Sigma-Aldrich (646563)
RNase inhibitor	50	Sigma-Aldrich (3335402001)
UltraPure Water	898	Invitrogen (10977049)

References

[1] M. M. Patnaik, S. A. Parikh, C. A. Hanson, and A. Tefferi, "Chronic myelomonocytic leukaemia: A concise clinical and pathophysiological review.," *British journal of haematology*, vol. 165, pp. 273–86, 3 May 2014.

[2] D. M. Berggren, M. Kjellander, E. Backlund, *et al.*, "Prognostic scoring systems and comorbidities in chronic myelomonocytic leukaemia: A nationwide population-based study," *Br J Haematol*, vol. 192, pp. 474–483, 3 2021. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pubmed/32501529`.

[3] S. Castano-Diez, M. Lopez-Guerra, C. Bosch-Castaneda, *et al.*, "Real-world data on chronic myelomonocytic leukemia: Clinical and molecular characteristics, treatment, emerging drugs, and patient outcomes," *Cancers (Basel)*, vol. 14, 17 2022. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pubmed/36077644`.

[4] A. Mekinian, E. Grignano, T. Braun, *et al.*, "Systemic inflammatory and autoimmune manifestations associated with myelodysplastic syndromes and chronic myelomonocytic leukaemia: A french multicentre retrospective study," *Rheumatology*, vol. 55, pp. 291–300, 2 Feb. 2016.

[5] M. F. Zahid, D. Barraco, T. L. Lasho, *et al.*, "Spectrum of autoimmune diseases and systemic inflammatory syndromes in patients with chronic myelomonocytic leukemia.," *Leukemia*
lymphoma, vol. 58, pp. 1488–1493, 6 Jun. 2017.

[6] A. T. Kuykendall, A. Hunter, L. Zhang, E. Padron, C. Talati, and R. S. Komrokji, "Bone marrow mastocytosis is independently associated with inferior survival in chronic myelomonocytic leukemia," *Blood*, vol. 134, pp. 2956–2956, Supplement$_1$ Nov. 2019.

[7] R. A. Mathew, J. M. Bennett, J. J. Liu, *et al.*, "Cutaneous manifestations in cmml: Indication of disease acceleration or transformation to aml and review of the literature.," *Leukemia research*, vol. 36, pp. 72–80, 1 Jan. 2012.

[8] M. M. Patnaik, E. A. Wassie, T. L. Lasho, C. A. Hanson, R. Ketterling, and A. Tefferi, "Blast transformation in chronic myelomonocytic leukemia: Risk factors, genetic features, survival, and treatment outcome," *American Journal of Hematology*, vol. 90, pp. 411–416, 5 May 2015.

[9] J. D. Khoury, E. Solary, O. Abla, *et al.*, "The 5th edition of the world health organization classification of haematolymphoid tumours: Myeloid and histiocytic/dendritic neoplasms," *Leukemia*, vol. 36, pp. 1703–1719, 7 Jul. 2022.

[10] E. Jaffe, N. Harris, H. Stein, and J. Vardiman, *World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. IARC Press, 2001.

[11] D. E. Rollison, N. Howlader, M. T. Smith, *et al.*, "Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the united states, 2001-2004, using data from the naaccr and seer programs," *Blood*, vol. 112, pp. 45–52, 1 Jul. 2008.

[12] A. G. Dinmohamed, Y. van Norden, O. Visser, *et al.*, "The use of medical claims to assess incidence, diagnostic procedures and initial treatment of myelodysplastic syndromes and chronic myelomonocytic leukemia in the netherlands," *Leukemia Research*, vol. 39, pp. 177–182, 2 Feb. 2015.

[13] E. Such, U. Germing, L. Malcovati, *et al.*, "Development and validation of a prognostic scoring system for patients with chronic myelomonocytic leukemia," *Blood*, vol. 121, pp. 3005–3015, 15 Apr. 2013.

[14] D. A. Arber, A. Orazi, R. P. Hasserjian, *et al.*, "International consensus classification of myeloid neoplasms and acute leukemias: Integrating morphologic, clinical, and genomic data," *Blood*, vol. 140, pp. 1200–1228, 11 Sep. 2022.

[15] G. Eisenwort, I. Sadovnik, A. Keller, *et al.*, "Phenotypic characterization of leukemia-initiating stem cells in chronic myelomonocytic leukemia," *Leukemia*, vol. 35, pp. 3176–3187, 11 Nov. 2021.

[16] R. Itzykson, O. Kosmider, A. Renneville, *et al.*, "Clonal architecture of chronic myelomonocytic leukemias," *Blood*, vol. 121, pp. 2186–2198, 12 Mar. 2013. [Online]. Available: `https://ashpublications.org/blood/article/121/12/2186/31473/Clonal-architecture-of-chronic-myelomonocytic`.

[17] M. C. Ferrall-Fairbanks, A. Dhawan, B. Johnson, *et al.*, "Progenitor hierarchy of chronic myelomonocytic leukemia identifies inflammatory monocytic-biased trajectory linked to worse outcomes," *Blood Cancer Discovery*, vol. 3, pp. 536–553, 6 Nov. 2022.

[18] A. S. Schnegg-Kaufmann, J. A. I. Thoms, G. S. Bhuyan, *et al.*, "Contribution of mutant hsc clones to immature and mature cells in mds and cmml, and variations with aza therapy," *Blood*, vol. 141, pp. 1316–1321, 11 Mar. 2023.

[19] P. S. Woll, U. Kjällquist, O. Chowdhury, *et al.*, "Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo," *Cancer Cell*, vol. 25, pp. 794–808, 6 Jun. 2014.

[20] N. Williams, J. Lee, E. Mitchell, *et al.*, "Life histories of myeloproliferative neoplasms inferred from phylogenies.," *Nature*, vol. 602, pp. 162–168, 7895 Feb. 2022.

[21] F. Blokzijl, J. de Ligt, M. Jager, *et al.*, "Tissue-specific mutation accumulation in human adult stem cells during life," *Nature*, vol. 538, pp. 260–264, 7624 Oct. 2016.

[22] M. A. Fabre, J. G. de Almeida, E. Fiorillo, *et al.*, "The longitudinal dynamics and natural history of clonal haematopoiesis," *Nature*, vol. 606, pp. 335–342, 7913 2022. [Online]. Available: `https://doi.org/10.1038/s41586-022-04785-z`.

[23] E. Mitchell, M. S. Chapman, N. Williams, *et al.*, "Clonal dynamics of haematopoiesis across the human lifespan," *Nature*, vol. 606, pp. 343–350, 7913 Jun. 2022.

[24] M. Xie, C. Lu, J. Wang, *et al.*, "Age-related mutations associated with clonal hematopoietic expansion and malignancies.," *Nature medicine*, vol. 20, pp. 1472–8, 12 Dec. 2014.

[25] S. Jaiswal, P. Fontanillas, J. Flannick, *et al.*, "Age-related clonal hematopoiesis associated with adverse outcomes," *New England Journal of Medicine*, vol. 371, pp. 2488–2498, 26 Dec. 2014.

[26] G. Genovese, A. K. Kähler, R. E. Handsaker, *et al.*, "Clonal hematopoiesis and blood-cancer risk inferred from blood dna sequence," *New England Journal of Medicine*, vol. 371, pp. 2477–2487, 26 2014. [Online]. Available: `https://www.nejm.org/doi/full/10.1056/NEJMoa1409405`.

[27] T. McKerrell, N. Park, T. Moreno, *et al.*, "Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis," *Cell Reports*, vol. 10, pp. 1239–1245, 8 Mar. 2015.

[28] A. L. Young, G. A. Challen, B. M. Birmann, and T. E. Druley, "Clonal haematopoiesis harbouring aml-associated mutations is ubiquitous in healthy adults," *Nature Communications*, vol. 7, p. 12 484, 1 Aug. 2016.

[29] K. Y. King, Y. Huang, D. Nakada, and M. A. Goodell, "Environmental influences on clonal hematopoiesis," *Experimental Hematology*, vol. 83, pp. 66–73, Mar. 2020.

[30] C. J. Watson, A. L. Papula, G. Y. P. Poon, *et al.*, "The evolutionary dynamics and fitness landscape of clonal hematopoiesis," *Science*, vol. 367, pp. 1449–1454, 6485 Mar. 2020.

[31] M. Gu, S. C. Kovilakam, W. G. Dunn, *et al.*, "Multiparameter prediction of myeloid neoplasia risk," *Nature Genetics*, vol. 55, pp. 1523–1530, 9 Sep. 2023.

[32] L. D. Weeks, A. Niroula, D. Neuberg, *et al.*, "Prediction of risk for myeloid malignancy in clonal hematopoiesis," *NEJM Evidence*, vol. 2, 5 Apr. 2023.

[33] C. C. Mason, J. S. Khorashad, S. K. Tantravahi, *et al.*, "Age-related mutations and chronic myelomonocytic leukemia," *Leukemia*, vol. 30, pp. 906–913, 4 Apr. 2016.

[34] B. J. Patel, B. Przychodzen, S. Thota, *et al.*, "Genomic determinants of chronic myelomonocytic leukemia," *Leukemia*, vol. 31, pp. 2815–2823, 12 2017.

[35] E. J. Baxter, L. M. Scott, P. J. Campbell, *et al.*, "Acquired mutation of the tyrosine kinase jak2 in human myeloproliferative disorders," *The Lancet*, vol. 365, pp. 1054–1061, 9464 Mar. 2005.

[36] S. Boettcher, P. G. Miller, R. Sharma, *et al.*, "A dominant-negative effect drives selection of *TP53* missense mutations in myeloid malignancies," *Science*, vol. 365, pp. 599–604, 6453 Aug. 2019.

[37] S. Jaiswal, P. Natarajan, A. J. Silver, *et al.*, "Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease.," *The New England journal of medicine*, vol. 377, pp. 111–121, 2 Jul. 2017.

[38] J. J. Fuster, S. MacLauchlan, M. A. Zuriaga, *et al.*, "Clonal hematopoiesis associated with tet2 deficiency accelerates atherosclerosis development in mice," *Science*, vol. 355, pp. 842–847, 6327 2017.

[39] T. P. Fidler, C. Xue, M. Yalcinkaya, *et al.*, "The aim2 inflammasome exacerbates atherosclerosis in clonal haematopoiesis," *Nature*, vol. 592, pp. 296–301, 7853 Apr. 2021.

[40] S. O. Abegunde, R. Buckstein, R. A. Wells, and M. J. Rauh, "An inflammatory environment containing tnf favors tet2-mutant clonal hematopoiesis," *Experimental Hematology*, vol. 59, pp. 60–65, 2018. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0301472X17308986`.

[41] Z. Cai, J. J. Kotzin, B. Ramdas, *et al.*, "Inhibition of inflammatory signaling in tet2 mutant preleukemic cells mitigates stress-induced abnormalities and clonal hematopoiesis," *Cell Stem Cell*, vol. 23, 833–849.e5, 6 Dec. 2018.

[42] M. Meisel, R. Hinterleitner, A. Pacis, *et al.*, "Microbial signals drive pre-leukaemic myeloproliferation in a tet2-deficient host," *Nature*, vol. 557, pp. 580–584, 7706 May 2018.

[43] D. Hormaechea-Agulla, K. A. Matatall, D. T. Le, *et al.*, "Chronic infection drives dnmt3a-loss-of-function clonal hematopoiesis via ifn signaling," *Cell Stem Cell*, vol. 28, 1428–1442.e6, 8 Aug. 2021.

[44] K. C. Higa, A. Goodspeed, J. S. Chavez, *et al.*, "Chronic interleukin-1 exposure triggers selection for *Cebpa* -knockout multipotent hematopoietic progenitors," *Journal of Experimental Medicine*, vol. 218, 6 Jun. 2021.

[45] S. Avagyan, J. E. Henninger, W. P. Mannherz, *et al.*, "Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis," *Science*, vol. 374, pp. 768–772, 6568 Nov. 2021.

[46] J. M. SanMiguel, E. Eudy, M. A. Loberg, *et al.*, "Distinct tumor necrosis factor alpha receptors dictate stem cell fitness versus lineage output in *Dnmt3a* -mutant clonal hematopoiesis," *Cancer Discovery*, vol. 12, pp. 2763–2773, 12 Dec. 2022.

[47] A. Rodriguez-Meira, R. Norfo, S. Wen, *et al.*, "Single-cell multi-omics identifies chronic inflammation as a driver of tp53-mutant leukemic evolution," *Nature Genetics*, vol. 55, pp. 1531–1541, 9 Sep. 2023.

[48] N. A. Jakobsen, S. Turkalj, A. G. Zeng, *et al.*, "Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging," *Cell Stem Cell*, vol. 31, 1127–1144.e17, 8 Aug. 2024.

[49] G. Coltro, A. A. Mangaonkar, T. L. Lasho, *et al.*, "Clinical, molecular, and prognostic correlates of number, type, and functional localization of tet2 mutations in chronic myelomonocytic leukemia (cmml)—a study of 1084 patients," *Leukemia*, vol. 34, pp. 1407–1421, 5 May 2020.

[50] J. G. Tate, S. Bamford, H. C. Jubb, *et al.*, "Cosmic: The catalogue of somatic mutations in cancer," *Nucleic Acids Research*, vol. 47, pp. D941–D947, D1 2018. [Online]. Available: `https://doi.org/10.1093/nar/gky1015`.

[51] C. A. Ortmann, D. G. Kent, J. Nangalia, *et al.*, "Effect of mutation order on myeloproliferative neoplasms," *New England Journal of Medicine*, vol. 372, pp. 601–612, 7 Feb. 2015.

[52] G. Schiroli, V. Kartha, F. M. Duarte, *et al.*, "Cell of origin epigenetic priming determines susceptibility to tet2 mutation," *Nature Communications*, vol. 15, p. 4325, 1 May 2024.

[53] C. Ricci, E. Fermo, S. Corti, *et al.*, "*RAS* mutations contribute to evolution of chronic myelomonocytic leukemia to the proliferative variant," *Clinical Cancer Research*, vol. 16, pp. 2246–2256, 8 Apr. 2010.

[54] R. M. Carr, D. Vorobyev, T. Lasho, *et al.*, "Ras mutations drive proliferative chronic myelomonocytic leukemia via a kmt2a-plk1 axis," *Nature Communications*, vol. 12, p. 2901, 1 2021. [Online]. Available: https://doi.org/10.1038/s41467-021-23186-w.

[55] M. M. Patnaik, P. A. Pophali, T. L. Lasho, *et al.*, "Clinical correlates, prognostic impact and survival outcomes in chronic myelomonocytic leukemia patients with the *JAK2* v617f mutation," *Haematologica*, vol. 104, e236–e239, 6 Jun. 2019.

[56] M. Leukemia, R. Itzykson, O. Kosmider, *et al.*, "Prognostic score including gene mutations in chronic myelomonocytic leukemia," *Journal of Clinical Oncology*, vol. 31, pp. 2428–2436, 19 Jul. 2013.

[57] M. M. Patnaik, R. Vallapureddy, T. L. Lasho, *et al.*, "Ezh2 mutations in chronic myelomonocytic leukemia cluster with asxl1 mutations and their co-occurrence is prognostically detrimental," *Blood Cancer Journal*, vol. 8, p. 12, 1 Jan. 2018.

[58] M. Meggendorfer, A. Roller, T. Haferlach, *et al.*, "Srsf2 mutations in 275 cases with chronic myelomonocytic leukemia (cmml)," *Blood*, vol. 120, pp. 3080–3088, 15 Oct. 2012.

[59] A. M. Taylor, J. Shih, G. Ha, *et al.*, "Genomic and functional approaches to understanding cancer aneuploidy," *Cancer Cell*, vol. 33, 676–689.e3, 4 Apr. 2018.

[60] G. Tang, L. Zhang, B. Fu, *et al.*, "Cytogenetic risk stratification of 417 patients with chronic myelomonocytic leukemia from a single institution," *American Journal of Hematology*, vol. 89, pp. 813–818, 8 Aug. 2014.

[61] E. A. Wassie, R. Itzykson, T. L. Lasho, *et al.*, "Molecular and prognostic correlates of cytogenetic abnormalities in chronic myelomonocytic leukemia: A mayo clinic- french consortium study," *American Journal of Hematology*, vol. 89, pp. 1111–1115, 12 Dec. 2014.

[62] U. Germing, A. Kündgen, and N. Gattermann, "Risk assessment in chronic myelomonocytic leukemia (cmml)," *Leukemia Lymphoma*, vol. 45, pp. 1311–1318, 7 Jul. 2004.

[63] C. Elena, A. Gallì, E. Such, *et al.*, "Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia," *Blood*, vol. 128, pp. 1408–1417, 10 Sep. 2016.

[64] F. Onida, H. M. Kantarjian, T. L. Smith, *et al.*, "Prognostic factors and scoring systems in chronic myelomonocytic leukemia: A retrospective analysis of 213 patients," *Blood*, vol. 99, pp. 840–849, 3 Feb. 2002.

[65] N. Gagelmann, A. Badbaran, D. W. Beelen, *et al.*, "A prognostic score including mutation profile and clinical features for patients with cmml undergoing stem cell transplantation," *Blood Advances*, vol. 5, pp. 1760–1769, 6 Mar. 2021.

[66] M. George, A. Smith, S. Sabesan, and G. Ranmuthugala, "Physical comorbidities and their relationship with cancer treatment and its outcomes in older adult populations: Systematic review," *JMIR Cancer*, vol. 7, e26425, 4 Oct. 2021.

[67] C. Franceschi, P. Garagnani, P. Parini, C. Giuliani, and A. Santoro, "Inflammaging: A new immune–metabolic viewpoint for age-related diseases," *Nature Reviews Endocrinology*, vol. 14, pp. 576–590, 10 Oct. 2018.

[68] E. Mejia-Ramirez and M. C. Florian, "Understanding intrinsic hematopoietic stem cell aging," *Haematologica*, vol. 105, pp. 22–37, 1 Jan. 2020.

[69] E. Frisan, J. Vandekerckhove, A. de Thonel, *et al.*, "Defective nuclear localization of hsp70 is associated with dyserythropoiesis and gata-1 cleavage in myelodysplastic syndromes," *Blood*, vol. 119, pp. 1532–1542, 6 Feb. 2012.

[70] B. L. Ebert, J. Pretz, J. Bosco, *et al.*, "Identification of rps14 as a 5q- syndrome gene by rna interference screen," *Nature*, vol. 451, pp. 335–339, 7176 Jan. 2008.

[71] H. Dolatshad, A. Pellagatti, F. G. Liberante, *et al.*, "Cryptic splicing events in the iron transporter abcb7 and other key target genes in sf3b1-mutant myelodysplastic syndromes," *Leukemia*, vol. 30, pp. 2322–2331, 12 Dec. 2016.

[72] V. Gelsi-Boyer, V. Trouplin, J. Roquain, *et al.*, "Asxl1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia," *British Journal of Haematology*, vol. 151, pp. 365–375, 4 Nov. 2010.

[73] M. M. Patnaik and A. Tefferi, "Chronic myelomonocytic leukemia: 2024 update on diagnosis, risk stratification and management," *American Journal of Hematology*, vol. 99, pp. 1142–1165, 6 Jun. 2024.

[74] M. M. Patnaik, D. Barraco, T. L. Lasho, *et al.*, "Dnmt3a mutations are associated with inferior overall and leukemia-free survival in chronic myelomonocytic leukemia," *American Journal of Hematology*, vol. 92, pp. 56–61, 1 Jan. 2017.

[75] L. Pleyer, M. Leisch, A. Kourakli, *et al.*, "Outcomes of patients with chronic myelomonocytic leukaemia treated with non-curative therapies: A retrospective cohort study," *The Lancet Haematology*, vol. 8, e135–e148, 2 Feb. 2021.

[76] T. Nösslinger, R. Reisner, H. Grüner, *et al.*, "Dysplastic versus proliferative cmml – a retrospective analysis of 91 patients from a single institution," *Leukemia Research*, vol. 25, pp. 741–747, 9 Sep. 2001.

[77] L. Ma, L. Jiang, W. Yang, *et al.*, "Real-world data of chronic myelomonocytic leukemia: A chinese single-center retrospective study," *Cancer Medicine*, vol. 10, pp. 1715–1725, 5 Mar. 2021.

[78] K. Usuki, S. Ohtake, S. Honda, *et al.*, "Real-world data of mds and cmml in japan: Results of jalsg clinical observational study-11 (jalsg-cs-11)," *International Journal of Hematology*, Dec. 2023.

[79] B. Xicoy, U. Germing, M. J. Jimenez, *et al.*, "Response to erythropoietic-stimulating agents in patients with chronic myelomonocytic leukemia," *Eur J Haematol*, vol. 97, pp. 33–38, 1 2016.

[80] P. Fenaux, U. Platzbecker, G. J. Mufti, *et al.*, "Luspatercept in patients with lower-risk myelodysplastic syndromes," *New England Journal of Medicine*, vol. 382, pp. 140–151, 2 Jan. 2020.

[81] U. Platzbecker, K. S. Götze, P. Kiewe, *et al.*, "Long-term efficacy and safety of luspatercept for anemia treatment in patients with lower-risk myelodysplastic syndromes: The phase ii pace-mds study," *Journal of Clinical Oncology*, vol. 40, pp. 3800–3807, 33 Nov. 2022.

[82] R. Komrokji, G. Garcia-Manero, L. Ades, *et al.*, "Sotatercept with long-term extension for the treatment of anaemia in patients with lower-risk myelodysplastic syndromes: A phase 2, dose-ranging trial," *The Lancet Haematology*, vol. 5, e63–e72, 2 Feb. 2018.

[83] F. Rabian, S. Chevret, B. Gruson, *et al.*, "Eltrombopag in chronic myelomonocytic leukemia with severe thrombocytopenia. a groupe francophone des myélodysplasies (gfm) study," *Leukemia*, Sep. 2024.

[84] E. Wattel, A. Guerci, B. Hecquet, *et al.*, "A randomized trial of hydroxyurea versus vp16 in adult chronic myelomonocytic leukemia. groupe français des myélodysplasies and european cmml group.," *Blood*, vol. 88, pp. 2480–7, 7 Oct. 1996.

[85] L. R. Silverman, E. P. Demakos, B. L. Peterson, *et al.*, "Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: A study of the cancer and leukemia group b," *Journal of Clinical Oncology*, vol. 20, pp. 2429–2440, 10 2002. [Online]. Available: https://ascopubs.org/doi/abs/10.1200/JCO.2002.04.117.

[86] P. Fenaux, G. J. Mufti, E. Hellstrom-Lindberg, *et al.*, "Articles effi cacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: A randomised, open-label, phase iii study," 2009. [Online]. Available: www.thelancet.com/.

[87] R. Itzykson, V. Santini, S. Thepot, *et al.*, "Decitabine versus hydroxyurea for advanced proliferative chronic myelomonocytic leukemia: Results of a randomized phase iii trial within the emsco network," *Journal of Clinical Oncology*, vol. 41, pp. 1888–1897, 10 2022. [Online]. Available: https://doi.org/10.1200/JCO.22.00437.

[88] M. M. Patnaik and A. Tefferi, "Chronic myelomonocytic leukemia: 2022 update on diagnosis, risk stratification, and management," *Am J Hematol*, vol. 97, pp. 352–372, 3 2022.

[89] G. Garcia-Manero, E. A. Griffiths, D. P. Steensma, *et al.*, "Oral cedazuridine/decitabine for mds and cmml: A phase 2 pharmacokinetic/pharmacodynamic randomized crossover study," *Blood*, vol. 136, pp. 674–683, 6 2020. [Online]. Available: https://doi.org/10.1182/blood.2019004143.

[90] J. Merlevede, N. Droin, T. Qin, *et al.*, "Mutation allele burden remains unchanged in chronic myelomonocytic leukaemia responding to hypomethylating agents," *Nature Communications*, vol. 7, Feb. 2016.

[91] E. Solary and R. Itzykson, "Chronic myelomonocytic leukemia gold jubilee chronic myelomonocytic leukemia," 2021. [Online]. Available: https://doi.org/10.3390/hemato2030026.

[92] M. Robin, L. C. de Wreede, E. Padron, *et al.*, "Role of allogeneic transplantation in chronic myelomonocytic leukemia: An international collaborative analysis," *Blood*, vol. 140, pp. 1408–1418, 12 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006497122007406.

[93] C. D. DiNardo, B. A. Jonas, V. Pullarkat, *et al.*, "Azacitidine and venetoclax in previously untreated acute myeloid leukemia," *New England Journal of Medicine*, vol. 383, pp. 617–629, 7 Aug. 2020.

[94] G. Montalban-Bravo, D. Hammond, C. D. DiNardo, *et al.*, "Activity of venetoclax-based therapy in chronic myelomonocytic leukemia," *Leukemia*, vol. 35, pp. 1494–1499, 5 May 2021.

[95] M. Sevin, F. Debeurme, L. Laplane, *et al.*, "Cytokine-like protein 1-induced survival of monocytes suggests a combined strategy targeting mcl1 and mapk in cmml," Tech. Rep., 2021. [Online]. Available: `http://ashpublications.org/blood/article-pdf/137/24/3390/1809947/bloodbld2020008729.pdf`.

[96] X. Wang, H. Hisha, T. Mizokami, *et al.*, "Mouse mesenchymal stem cells can support human hematopoiesis both in vitro and in vivo: The crucial role of neural cell adhesion molecule," *Haematologica*, vol. 95, pp. 884–891, 6 Jun. 2010.

[97] E. Padron, J. S. Painter, S. Kunigal, *et al.*, "Gm-csf–dependent pstat5 sensitivity is a feature with therapeutic potential in chronic myelomonocytic leukemia," *Blood*, vol. 121, pp. 5068–5077, 25 Jun. 2013.

[98] A. M. Hunter, H. Newman, A. E. Dezern, *et al.*, "Integrated human and murine clinical study establishes clinical efficacy of ruxolitinib in chronic myelomonocytic leukemia," *Clinical Cancer Research*, vol. 27, pp. 6095–6105, 22 Nov. 2021.

[99] A. Yoshimi, M. E. Balasis, A. Vedder, *et al.*, "Robust patient-derived xenografts of mds/mpn overlap syndromes capture the unique characteristics of cmml and jmml," *Blood*, vol. 130, pp. 397–407, 4 Jul. 2017.

[100] Y. Ma, L. R. Rix, Q. Zhang, *et al.*, "Pacritinib (pac) synergistically potentiates azacitidine (5aza) cytotoxicity in chronic myelomonocytic leukemia (cmml)," *Blood*, vol. 126, pp. 1658–1658, 23 Dec. 2015.

[101] M. M. Patnaik, D. A. Sallman, A. A. Mangaonkar, *et al.*, "Phase 1 study of lenzilumab, a recombinant anti-human gm-csf antibody, for chronic myelomonocytic leukemia," *Blood*, vol. 136, pp. 909–913, 7 2020. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pubmed/32294158`.

[102] D. Hiwase, D. Ross, D. Yeung, *et al.*, "Durable efficacy of lenzilumab plus azacitidine in newly diagnosed proliferative cmml patients: Interim analysis after four years of study commencement," *Blood*, vol. 146, no. Supplement 1, pp. 3855–3855, Nov. 2025. eprint: `https://ashpublications.org/blood/article-pdf/146/Supplement1/3855/2449731/blood-6602-main.pdf`. [Online]. Available: `https://doi.org/10.1182/blood-2025-3855`.

[103] M. E. Bahar, H. J. Kim, and D. R. Kim, "Targeting the ras/raf/mapk pathway for cancer therapy: From mechanism to clinical studies," *Signal Transduction and Targeted Therapy*, vol. 8, p. 455, 1 Dec. 2023.

[104] A. B. Patel, A. Osman, T. D. Pomicter, S. K. Tantravahi, and M. W. Deininger, "Concerto (nct04409639): A phase 2 trial of cobimetinib in newly diagnosed and hma-treated cmml patients with ras pathway mutations," *Blood*, vol. 140, pp. 4060–4061, Supplement 1 Nov. 2022.

[105] J. Wang, X. Yao, and J. Huang, "New tricks for human farnesyltransferase inhibitor: Cancer and beyond," *MedChemComm*, vol. 8, pp. 841–854, 5 2017.

[106] M. M. Patnaik, M. A. Sekeres, A. DeZern, *et al.*, "854p final results of a phase ii study of tipifarnib in chronic myelomonocytic leukemia (cmml) and other myelodysplastic/myeloproliferative neoplasms (mds/mpn)," *Annals of Oncology*, vol. 32, S783–S784, Sep. 2021. [Online]. Available: https://doi.org/10.1016/j.annonc.2021.08.147.

[107] L. Cimmino, I. Dolgalev, Y. Wang, *et al.*, "Restoration of tet2 function blocks aberrant self-renewal and leukemia progression," *Cell*, vol. 170, 1079–1095.e20, 6 Sep. 2017.

[108] M. Deininger, J. Radich, T. C. Burn, R. Huber, D. Paranagama, and S. Verstovsek, "The effect of long-term ruxolitinib treatment on jak2p.v617f allele burden in patients with myelofibrosis," *Blood*, vol. 126, pp. 1551–1554, 13 Sep. 2015.

[109] S. Niyongere, N. Lucas, J. M. Zhou, *et al.*, "Heterogeneous expression of cytokines accounts for clinical diversity and refines prognostication in cmml," *Leukemia*, vol. 33, pp. 205–216, 1 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/30026572.

[110] A. Franzini, A. D. Pomicter, D. Yan, *et al.*, "The transcriptome of cmml monocytes is highly inflammatory and reflects leukemia-specific and age-related alterations," *Blood Adv*, vol. 3, pp. 2949–2961, 20 2019.

[111] P. Deschamps, M. Wacheux, A. Gosseye, *et al.*, "Cxcl8 secreted by immature granulocytes inhibits wt hematopoiesis in chronic myelomonocytic leukemia," *Journal of Clinical Investigation*, vol. 134, 22 Nov. 2024.

[112] E. Pronier, A. Imanci, D. Selimoglu-Buet, *et al.*, "Macrophage migration inhibitory factor is overproduced through egr1 in tet2 low resting monocytes," *Communications Biology*, vol. 5, 1 Dec. 2022.

[113] N. Droin, A. Jacquel, J.-B. Hendra, *et al.*, "Alpha-defensins secreted by dysplastic granulocytes inhibit the differentiation of monocytes in chronic myelomonocytic leukemia," *Blood*, vol. 115, pp. 78–88, 1 Jan. 2010.

[114] K. Geissler, L. Ohler, M. Födinger, *et al.*, "Interleukin 10 inhibits growth and granulocyte/macrophage colony-stimulating factor production in chronic myelomonocytic leukemia cells," *J Exp Med*, vol. 184, pp. 1377–1384, 4 1996.

[115] M. Tahiliani, K. P. Koh, Y. Shen, *et al.*, "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian dna by mll partner tet1," *Science*, vol. 324, pp. 930–935, 5929 May 2009.

[116] S. Ito, L. Shen, Q. Dai, *et al.*, "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine," *Science*, vol. 333, pp. 1300–1303, 6047 2011.

[117] Y. F. He, B. Z. Li, Z. Li, *et al.*, "Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian dna," *Science*, vol. 333, pp. 1303–1307, 6047 2011.

[118] J. U. Guo, Y. Su, C. Zhong, G.-l. Ming, and H. Song, "Hydroxylation of 5-methylcytosine by tet1 promotes active dna demethylation in the adult brain," *Cell*, vol. 145, pp. 423–434, 3 Apr. 2011.

[119] R. J. Boorstein, C. Archie, D. R. Marenstein, *et al.*, "Definitive identification of mammalian 5-hydroxymethyluracil dna n-glycosylase activity as smug1," *Journal of Biological Chemistry*, vol. 276, pp. 41 991–41 997, 45 Nov. 2001.

[120] V. Valinluck and L. C. Sowers, "Endogenous cytosine damage products alter the site selectivity of human dna maintenance methyltransferase dnmt1," *Cancer Research*, vol. 67, pp. 946–950, 3 Feb. 2007.

[121] E. Li, T. H. Bestor, and R. Jaenisch, "Targeted mutation of the dna methyltransferase gene results in embryonic lethality," *Cell*, vol. 69, pp. 915–926, 6 Jun. 1992.

[122] M. Okano, D. W. Bell, D. A. Haber, and E. Li, "Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development," *Cell*, vol. 99, pp. 247–257, 3 Oct. 1999.

[123] M. Ehrlich, M. A. Gama-Sosa, L.-H. Huang, *et al.*, "Amount and distribution of 5-methylcytosine in human dna from different types of tissues or cells," *Nucleic Acids Research*, vol. 10, pp. 2709–2721, 8 1982.

[124] R. Lister, M. Pelizzola, R. H. Dowen, *et al.*, "Human dna methylomes at base resolution show widespread epigenomic differences," *Nature*, vol. 462, pp. 315–322, 7271 Nov. 2009.

[125] S. Saxonov, P. Berg, and D. L. Brutlag, "A genome-wide analysis of cpg dinucleotides in the human genome distinguishes two distinct classes of promoters," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 1412–1417, 5 Jan. 2006.

[126] R. Illingworth, A. Kerr, D. DeSousa, *et al.*, "A novel cpg island set identifies tissue-specific methylation at developmental gene loci," *PLoS Biology*, vol. 6, e22, 1 Jan. 2008.

[127] J. S. K. Bell and P. M. Vertino, "Orphan cpg islands define a novel class of highly active enhancers," *Epigenetics*, vol. 12, pp. 449–464, 6 Jun. 2017.

[128] T. Pachano, V. Sánchez-Gaya, T. Ealo, *et al.*, "Orphan cpg islands amplify poised enhancer regulatory activity and determine target gene responsiveness," *Nature Genetics*, vol. 53, pp. 1036–1049, 7 Jul. 2021.

[129] T. Mohandas, R. S. Sparkes, and L. J. Shapiro, "Reactivation of an inactive human x chromosome: Evidence for x inactivation by dna methylation," *Science*, vol. 211, pp. 393–396, 4480 Jan. 1981.

[130] S. Eden, M. Constancia, T. Hashimshony, *et al.*, "An upstream repressor element plays a role in *Igf2* imprinting," *The EMBO Journal*, vol. 20, pp. 3518–3525, 13 Jul. 2001. [Online]. Available: `https://doi.org/10.1093/emboj/20.13.3518`.

[131] J. D. Hollister and B. S. Gaut, "Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression," *Genome Research*, vol. 19, pp. 1419–1428, 8 Aug. 2009.

[132] A. Bird, "Dna methylation patterns and epigenetic memory," *Genes Development*, vol. 16, pp. 6–21, 1 Jan. 2002.

[133] R. Greenfield, A. Tabib, I. Keshet, *et al.*, "Role of transcription complexes in the formation of the basal methylation pattern in early development," *Proceedings of the National Academy of Sciences*, vol. 115, pp. 10 387–10 391, 41 Oct. 2018.

[134] J. L. Huppert and S. Balasubramanian, "G-quadruplexes in promoters throughout the human genome," *Nucleic Acids Research*, vol. 35, pp. 406–413, 2 Jan. 2007.

[135] R. Halder, K. Halder, P. Sharma, G. Garg, S. Sengupta, and S. Chowdhury, "Guanine quadruplex dna structure restricts methylation of cpg dinucleotides genome-wide," *Molecular BioSystems*, vol. 6, p. 2439, 12 2010.

[136] F. Larsen, G. Gundersen, R. Lopez, and H. Prydz, "Cpg islands as gene markers in the human genome," *Genomics*, vol. 13, pp. 1095–1107, 4 Aug. 1992.

[137] F. Antequera, "Structure, function and evolution of cpg island promoters," *Cellular and Molecular Life Sciences (CMLS)*, vol. 60, pp. 1647–1658, 8 Aug. 2003.

[138] R. Edgar, P. P. C. Tan, E. Portales-Casamar, and P. Pavlidis, "Meta-analysis of human methylomes reveals stably methylated sequences surrounding cpg islands associated with high gene expression," *Epigenetics Chromatin*, vol. 7, p. 28, 1 Dec. 2014.

[139] N. Loyfer, J. Magenheim, A. Peretz, *et al.*, "A dna methylation atlas of normal human cell types," *Nature*, vol. 613, pp. 355–364, 7943 Jan. 2023.

[140] J. Ernst and M. Kellis, "Chromhmm: Automating chromatin-state discovery and characterization," *Nature Methods*, vol. 9, pp. 215–216, 3 Mar. 2012.

[141] X. S. Liu, H. Wu, X. Ji, *et al.*, "Editing dna methylation in the mammalian genome," *Cell*, vol. 167, 233–247.e17, 1 Sep. 2016.

[142] S. Morita, H. Noguchi, T. Horii, *et al.*, "Targeted dna demethylation in vivo using dcas9–peptide repeat and scfv–tet1 catalytic domain fusions," *Nature Biotechnology*, vol. 34, pp. 1060–1065, 10 Oct. 2016.

[143] P. Stepper, G. Kungulovski, R. Z. Jurkowska, *et al.*, "Efficient targeted dna methylation with chimeric dcas9–dnmt3a–dnmt3l methyltransferase," *Nucleic Acids Research*, vol. 45, pp. 1703–1713, 4 Feb. 2017.

[144] I. K. Mann, R. Chatterjee, J. Zhao, *et al.*, "Cg methylated microarrays identify a novel methylated sequence bound by the cebpb|atf4 heterodimer that is active in vivo," *Genome Research*, vol. 23, pp. 988–997, 6 Jun. 2013.

[145] K. Gaston and M. Fried, "Cpg methylation has differential effects on the binding of yy1 and ets proteins to the bi-directional promoter of the surf-1 and surf-2 genes," *Nucleic Acids Research*, vol. 23, pp. 901–909, 6 1995.

[146] Y. Yin, E. Morgunova, A. Jolma, *et al.*, "Impact of cytosine methylation on dna binding specificities of human transcription factors," *Science*, vol. 356, 6337 May 2017.

[147] R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, and A. P. Bird, "Identification of a mammalian protein that binds specifically to dna containing methylated cpgs," *Cell*, vol. 58, pp. 499–507, 3 Aug. 1989.

[148] F. Antequera, "<scp>cpg</scp> islands and <scp>dna</scp> methylation," in *Encyclopedia of Life Sciences*. Wiley, Apr. 2017, pp. 1–7.

[149] L. M. Iyer, S. Abhiman, and L. Aravind, "Natural history of eukaryotic dna methylation systems," in 2011, pp. 25–104.

[150] M. Ko, J. An, H. S. Bandukwala, *et al.*, "Modulation of tet2 expression and 5-methylcytosine oxidation by the cxxc domain protein idax," *Nature*, vol. 497, pp. 122–126, 7447 May 2013.

[151] C. Xu, K. Liu, M. Lei, *et al.*, "Dna sequence recognition of human cxxc domains and their structural determinants," *Structure*, vol. 26, 85–95.e3, 1 Jan. 2018.

[152] K. Williams, J. Christensen, M. T. Pedersen, *et al.*, "Tet1 and hydroxymethylcytosine in transcription and dna methylation fidelity," *Nature*, vol. 473, pp. 343–348, 7347 May 2011.

[153] S.-G. Jin, Z.-M. Zhang, T. L. Dunwell, *et al.*, "Tet3 reads 5-carboxylcytosine through its cxxc domain and is a potential guardian against neurodegeneration," *Cell Reports*, vol. 14, pp. 493–505, 3 Jan. 2016.

[154] L. Wang, P. A. Ozark, E. R. Smith, *et al.*, "Tet2 coactivates gene expression through demethylation of enhancers," *Science Advances*, vol. 4, 11 Nov. 2018.

[155] K. D. Rasmussen, G. Jia, J. V. Johansen, *et al.*, "Loss of *TET2* in hematopoietic cells leads to dna hypermethylation of active enhancers and induction of leukemogenesis," *Genes Development*, vol. 29, pp. 910–922, 9 May 2015.

[156] K. D. Rasmussen, I. Berest, S. Keler, *et al.*, "Tet2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells," *Genome Research*, vol. 29, pp. 564–575, 4 Apr. 2019.

[157] R. Broome, I. Chernukhin, S. Jamieson, *et al.*, "Tet2 is a component of the estrogen receptor complex and controls 5mc to 5hmc conversion at estrogen receptor cis-regulatory regions," *Cell Reports*, vol. 34, p. 108 776, 8 Feb. 2021.

[158] Y. Wang, M. Xiao, X. Chen, *et al.*, "Wt1 recruits tet2 to regulate its target gene expression and suppress leukemia cell proliferation," *Molecular Cell*, vol. 57, pp. 662–673, 4 Feb. 2015.

[159] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, and A. A. Alizadeh, "Profiling tumor infiltrating immune cells with cibersort," in 2018, pp. 243–259.

[160] R. T. Pop, A. Pisante, D. Nagy, *et al.*, "Identification of mammalian transcription factors that bind to inaccessible chromatin," *Nucleic Acids Research*, vol. 51, pp. 8480–8495, 16 Sep. 2023.

[161] Q. Chen, Y. Chen, C. Bian, R. Fujiki, and X. Yu, "Tet2 promotes histone o-glcnacylation during gene transcription," *Nature*, vol. 493, pp. 561–564, 7433 Jan. 2013.

[162] J. Zhang, Y. K. Lieu, A. M. Ali, *et al.*, "Disease-associated mutation in *SRSF2* misregulates splicing by altering rna-binding affinities," *Proceedings of the National Academy of Sciences*, vol. 112, 34 Aug. 2015.

[163] M. Uhlén, L. Fagerberg, B. M. Hallström, *et al.*, "Tissue-based map of the human proteome," *Science*, vol. 347, 6220 Jan. 2015.

[164] C. Quivoron, L. Couronné, V. D. Valle, *et al.*, "Tet2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis," *Cancer Cell*, vol. 20, pp. 25–38, 1 2011.

[165] M. Ko, H. S. Bandukwala, J. An, *et al.*, "Ten-eleven-translocation 2 (tet2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 14 566–14 571, 35 Aug. 2011.

[166] K. Moran-Crusio, L. Reavie, A. Shih, *et al.*, "Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation," *Cancer Cell*, vol. 20, pp. 11–24, 1 2011.

[167] Z. Li, X. Cai, C. L. Cai, *et al.*, "Deletion of tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies," in *Blood*, vol. 118, American Society of Hematology, Oct. 2011, pp. 4509–4518.

[168] K. Shide, T. Kameda, H. Shimoda, *et al.*, "Tet2 is essential for survival and hematopoietic stem cell homeostasis," *Leukemia*, vol. 26, pp. 2216–2223, 10 Oct. 2012.

[169] E. Pronier, C. Almire, H. Mokrani, *et al.*, "Inhibition of tet2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors," *Blood*, vol. 118, pp. 2551–2555, 9 Sep. 2011.

[170] H. Kunimoto, A. S. McKenney, C. Meydan, *et al.*, "Aid is a key regulator of myeloid/erythroid differentiation and dna methylation in hematopoietic stem/progenitor cells," *Blood*, vol. 129, pp. 1779–1790, 13 Mar. 2017.

[171] K. Ito, J. Lee, S. Chrysanthou, *et al.*, "Non-catalytic roles of tet2 are essential to regulate hematopoietic stem and progenitor cell homeostasis," *Cell Reports*, vol. 28, 2480–2490.e4, 10 Sep. 2019.

[172] M. Farlik, F. Halbritter, F. Müller, *et al.*, "Dna methylation dynamics of human hematopoietic stem cell differentiation," *Cell Stem Cell*, vol. 19, pp. 808–822, 6 Dec. 2016.

[173] M. R. Corces, J. D. Buenrostro, B. Wu, *et al.*, "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution," *Nat Genet*, vol. 48, pp. 1193–1203, 10 2016.

[174] J. D. Buenrostro, M. R. Corces, C. A. Lareau, *et al.*, "Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation," *Cell*, vol. 173, 1535–1548.e16, 6 May 2018.

[175] N. Novershtern, A. Subramanian, L. N. Lawton, *et al.*, "Densely interconnected transcriptional circuits control cell states in human hematopoiesis," *Cell*, vol. 144, pp. 296–309, 2 Jan. 2011.

[176] L. Chen, M. Kostadima, J. H. A. Martens, *et al.*, "Transcriptional diversity during lineage commitment of human blood progenitors," *Science*, vol. 345, 6204 Sep. 2014.

[177] V. W. Yu, R. Z. Yusuf, T. Oki, *et al.*, "Epigenetic memory underlies cell-autonomous heterogeneous behavior of hematopoietic stem cells," *Cell*, vol. 167, 1310–1322.e17, 5 Nov. 2016.

[178] L. Biasco, D. Pellin, S. Scala, *et al.*, "In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases," *Cell Stem Cell*, vol. 19, pp. 107–119, 1 Jul. 2016.

[179] Y. Nakauchi, A. Azizi, D. Thomas, *et al.*, "The cell type–specific 5hmc landscape and dynamics of healthy human hematopoiesis and *TET2* -mutant preleukemia," *Blood Cancer Discovery*, vol. 3, pp. 346–367, 4 Jul. 2022. [Online]. Available: https://aacrjournals.org/bloodcancerdiscov/article/3/4/346/705269/The-Cell-Type-Specific-5hmC-Landscape-and-Dynamics.

[180] M. Tulstrup, M. Soerensen, J. W. Hansen, *et al.*, "Tet2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis," *Nature Communications*, vol. 12, p. 6061, 1 2021. [Online]. Available: https://doi.org/10.1038/s41467-021-26093-2.

[181] A. H. Cull, B. Snetsinger, R. Buckstein, R. A. Wells, and M. J. Rauh, "Tet2 restrains inflammatory gene expression in macrophages," *Experimental Hematology*, vol. 55, 56–70.e13, Nov. 2017.

[182] S. Sano, K. Oshima, Y. Wang, *et al.*, "Tet2-mediated clonal hematopoiesis accelerates heart failure through a mechanism involving the il-1/nlrp3 inflammasome," *Journal of the American College of Cardiology*, vol. 71, pp. 875–886, 8 Feb. 2018.

[183] W. Pan, S. Zhu, K. Qu, *et al.*, "The dna methylcytosine dioxygenase tet2 sustains immunosuppressive function of tumor-infiltrating myeloid cells to promote melanoma progression," *Immunity*, vol. 47, 284–297.e5, 2 Aug. 2017.

[184] F. Caiado, L. V. Kovtonyuk, N. G. Gonullu, J. Fullin, S. Boettcher, and M. G. Manz, "Aging drives tet2+/- clonal hematopoiesis via il-1 signaling," *Blood*, 2022. [Online]. Available: https://doi.org/10.1182/blood.2022016835.

[185] S. S. Burns, R. Kumar, S. K. Pasupuleti, K. So, C. Zhang, and R. Kapur, "Il-1r1 drives leukemogenesis induced by tet2 loss," *Leukemia*, vol. 36, pp. 2531–2534, 10 2022. [Online]. Available: https://doi.org/10.1038/s41375-022-01665-3.

[186] T. Muto, C. S. Walker, K. Choi, *et al.*, "Adaptive response to inflammation contributes to sustained myelopoiesis and confers a competitive advantage in myelodysplastic syndrome hscs," *Nature Immunology*, vol. 21, pp. 535–545, 5 May 2020.

[187] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, pp. 646–674, 5 Mar. 2011. [Online]. Available: https://doi.org/10.1016/j.cell.2011.02.013.

[188] E. Cerami, J. Gao, U. Dogrusoz, *et al.*, "The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data," *Cancer Discovery*, vol. 2, pp. 401–404, 5 May 2012.

[189] J. C. Chang, G. F. Temple, R. F. Trecartin, and Y. W. Kan, "Suppression of the nonsense mutation in homozygous 0 thalassaemia," *Nature*, vol. 281, pp. 602–603, 5732 Oct. 1979.

[190] M. Ko, Y. Huang, A. M. Jankowska, *et al.*, "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant tet2," *Nature*, vol. 468, pp. 839–843, 7325 Dec. 2010.

[191] F. André, M. Arnedos, A. S. Baras, *et al.*, "Aacr project genie: Powering precision medicine through an international consortium," *Cancer Discovery*, vol. 7, pp. 818–831, 8 Aug. 2017.

[192] C. B. Jing, C. Fu, N. Prutsch, M. Wang, S. He, and A. T. Look, "Synthetic lethal targeting of tet2-mutant hematopoietic stem and progenitor cells (hspcs) with top1-targeted drugs and parp1 inhibitors," *Leukemia*, vol. 34, pp. 2992–3006, 11 Nov. 2020.

[193] S. Maifrede, B. V. Le, M. Nieborowska-Skorska, *et al.*, "Tet2 and dnmt3a mutations exert divergent effects on dna repair and sensitivity of leukemia cells to parp inhibitors," *Cancer Research*, vol. 81, pp. 5089–5101, 19 Oct. 2021.

[194] F. Pan, T. S. Wingo, Z. Zhao, *et al.*, "Tet2 loss leads to hypermutagenicity in haematopoietic stem/progenitor cells," *Nature Communications*, vol. 8, 2017.

[195] H. Awada, Y. Nagata, A. Goyal, *et al.*, "Invariant phenotype and molecular association of biallelic tet2 mutant myeloid neoplasia," *Blood Advances*, vol. 3, pp. 339–349, 3 Feb. 2019.

[196] F. Delhommeau, S. Dupont, V. D. Valle, *et al.*, "Mutation in tet2 in myeloid cancers a bs tr ac t," Tech. Rep., 2009, pp. 2289–301.

[197] M. M. Patnaik, M. F. Zahid, T. L. Lasho, *et al.*, "Number and type of tet2 mutations in chronic myelomonocytic leukemia and their clinical relevance," *Blood Cancer Journal*, vol. 6, e472–e472, 9 Sep. 2016.

[198] L. Palomo, M. Meggendorfer, S. Hutter, *et al.*, "Molecular landscape and clonal architecture of adult myelodysplastic/myeloproliferative neoplasms," *Blood*, vol. 136, pp. 1851–1862, 16 Oct. 2020.

[199] A. M. Jankowska, H. Makishima, R. V. Tiu, *et al.*, "Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: Utx„ezh2,and dnmt3a," *Blood*, vol. 118, pp. 3932–3941, 14 Oct. 2011.

[200] N. Garcia-Gisbert, L. Arenillas, D. Roman-Bravo, *et al.*, "Multi-hit tet2 mutations as a differential molecular signature of oligomonocytic and overt chronic myelomonocytic leukemia," *Leukemia*, vol. 36, pp. 2922–2926, 12 Dec. 2022.

[201] C. Pérez, N. Martínez-Calle, J. I. Martín-Subero, *et al.*, "Tet2 mutations are associated with specific 5-methylcytosine and 5-hydroxymethylcytosine profiles in patients with chronic myelomonocytic leukemia," *PLoS ONE*, vol. 7, 2 Feb. 2012.

[202] J. Yamazaki, R. Taby, A. Vasanthakumar, *et al.*, "Efects of tet2 mutations on dna methylation in chronic myelomonocytic leukemia," *Epigenetics*, vol. 7, pp. 201–207, 2 2012.

[203] L. Palomo, R. Malinverni, M. Cabezón, *et al.*, "Dna methylation profile in chronic myelomonocytic leukemia associates with distinct clinical, biological and genetic features," *Epigenetics*, vol. 13, pp. 8–18, 1 Jan. 2018.

[204] K. Meldi, T. Qin, F. Buchi, *et al.*, "Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia," *Journal of Clinical Investigation*, vol. 125, pp. 1857–1872, 5 May 2015.

[205] J. L. Berg, B. Perfler, S. Hatzl, *et al.*, "Micro-rna-125a mediates the effects of hypomethylating agents in chronic myelomonocytic leukemia," *Clinical Epigenetics*, vol. 13, p. 1, 1 Dec. 2021.

[206] R. Aucagne, N. Droin, J. Paggetti, *et al.*, "Transcription intermediary factor 1 is a tumor suppressor in mouse and human chronic myelomonocytic leukemia," *Journal of Clinical Investigation*, vol. 121, pp. 2361–2370, 6 Jun. 2011.

[207] D. Selimoglu-Buet, J. Rivière, H. Ghamlouch, *et al.*, "A mir-150/tet3 pathway regulates the generation of mouse and human non-classical monocyte subset," *Nature Communications*, vol. 9, p. 5455, 1 Dec. 2018.

[208] S. Paget, "The distribution of secondary growths in cancer of the breast. 1889.," *Cancer metastasis reviews*, vol. 8, pp. 98–101, 2 Aug. 1989.

[209] A. Hamblin, A. Burns, C. Tham, *et al.*, "Development and evaluation of the clinical utility of a next generation sequencing (ngs) tool for myeloid disorders," *Blood*, vol. 124, pp. 2373–2373, 21 Dec. 2014.

[210] G. A. V. der Auwera, M. O. Carneiro, C. Hartl, *et al.*, "From fastq data to high-confidence variant calls: The genome analysis toolkit best practices pipeline," *Current Protocols in Bioinformatics*, vol. 43, 1 Oct. 2013.

[211] O. Chowdhury, "Characterisation and targeting of stem cells in myelodysplastic syndromes," Ph.D. dissertation, University of Oxford, 2013.

[212] M. Dimitriou, P. S. Woll, T. Mortera-Blanco, *et al.*, "Perturbed hematopoietic stem and progenitor cell hierarchy in myelodysplastic syndromes patients with monosomy 7 as the sole cytogenetic abnormality," *Oncotarget*, vol. 7, pp. 72 685–72 698, 45 Nov. 2016.

[213] M. Hagemann-Jensen, C. Ziegenhain, P. Chen, *et al.*, "Single-cell rna counting at allele and isoform resolution using smart-seq3," *Nature Biotechnology*, vol. 38, pp. 708–714, 6 Jun. 2020.

[214] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, "Atac-seq: A method for assaying chromatin accessibility genome-wide," *Curr Protoc Mol Biol*, vol. 109, pp. 21.29.1–21.29.9, 2015.

[215] F. Krueger, F. James, P. Ewels, *et al.*, *Felixkrueger/trimgalore: V0.6.7*, 2021. [Online]. Available: `https://github.com/FelixKrueger/TrimGalore`.

[216] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with bowtie 2," *Nat Methods*, vol. 9, pp. 357–359, 4 2012.

[217] P. Danecek, J. K. Bonfield, J. Liddle, *et al.*, "Twelve years of samtools and bcftools," *GigaScience*, vol. 10, giab008, 2 2021. [Online]. Available: `https://doi.org/10.1093/gigascience/giab008`.

[218] B. Institute, "Picard toolkit," *Broad Institute, GitHub repository*, 2019. [Online]. Available: `https://broadinstitute.github.io/picard/`.

[219] J. Ou, H. Liu, J. Yu, *et al.*, "Atacseqqc: A bioconductor package for post-alignment quality assessment of atac-seq data," *BMC Genomics*, vol. 19, p. 169, 1 2018. [Online]. Available: `https://doi.org/10.1186/s12864-018-4559-3`.

[220] Y. Zhang, T. Liu, C. A. Meyer, *et al.*, "Model-based analysis of chip-seq (macs)," *Genome Biology*, vol. 9, R137, 9 2008. [Online]. Available: `https://doi.org/10.1186/gb-2008-9-9-r137`.

[221] G. Yu, L.-G. Wang, and Q.-Y. He, "Chipseeker: An r/bioconductor package for chip peak annotation, comparison and visualization," *Bioinformatics*, vol. 31, pp. 2382–2383, 14 Jul. 2015.

[222] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for rna-seq data with deseq2," *Genome Biology*, vol. 15, p. 550, 12 2014. [Online]. Available: `https://doi.org/10.1186/s13059-014-0550-8`.

[223] A. N. Schep, B. Wu, J. D. Buenrostro, and W. J. Greenleaf, "Chromvar: Inferring transcription-factor-associated accessibility from single-cell epigenomic data," *Nature Methods*, vol. 14, pp. 975–978, 10 Oct. 2017.

[224] D. Baranasic, *Jaspar2020: Data package for jaspar database (version 2022)*, 2022. [Online]. Available: `http://jaspar.genereg.net/`.

[225] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, 2016.

[226] Z. Gu, R. Eils, and M. Schlesner, "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," *Bioinformatics*, vol. 32, pp. 2847–2849, 18 Sep. 2016.

[227] T. Smith, A. Heger, and I. Sudbery, "Umi-tools: Modeling sequencing errors in unique molecular identifiers to improve quantification accuracy," *Genome Res*, vol. 27, pp. 491–499, 3 2017. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pubmed/28100584`.

[228] A. Dobin, C. A. Davis, F. Schlesinger, *et al.*, "Star: Ultrafast universal rna-seq aligner," *Bioinformatics*, vol. 29, pp. 15–21, 1 2013. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pubmed/23104886`.

[229] Y. Liao, G. K. Smyth, and W. Shi, "Featurecounts: An efficient general purpose program for assigning sequence reads to genomic features," *Bioinformatics*, vol. 30, pp. 923–930, 7 2014.

[230] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "Multiqc: Summarize analysis results for multiple tools and samples in a single report.," *Bioinformatics (Oxford, England)*, vol. 32, pp. 3047–8, 19 Oct. 2016.

[231] M. E. Ritchie, B. Phipson, D. Wu, *et al.*, "Limma powers differential expression analyses for rna-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, e47–e47, 7 Apr. 2015.

[232] T. Wu, E. Hu, S. Xu, *et al.*, "Clusterprofiler 4.0: A universal enrichment tool for interpreting omics data," *Innovation (Camb)*, vol. 2, p. 100 141, 3 2021. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pubmed/34557778`.

[233] Gennady, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, and A. K. Sergushichev, "Fast gene set enrichment analysis," *bioRxiv*, 2021.

[234] T. M. Therneau, *A package for survival analysis in r*, 2024.

[235] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet.journal*, vol. 17, p. 10, 1 May 2011.

[236] S. J. Fleming, M. D. Chaffin, A. Arduini, *et al.*, "Unsupervised removal of systematic background noise from droplet-based single-cell experiments using cellbender," *Nature Methods*, vol. 20, pp. 1323–1335, 9 Sep. 2023.

[237] D. Neavin, A. Senabouth, H. Arora, *et al.*, "Demuxafy: Improvement in droplet assignment by integrating multiple single-cell demultiplexing and doublet detection methods," *Genome Biology*, vol. 25, p. 94, 1 Apr. 2024.

[238] H. Heaton, A. M. Talman, A. Knights, *et al.*, "Souporcell: Robust clustering of single-cell rna-seq data by genotype without reference genotypes," *Nature Methods*, vol. 17, pp. 615–620, 6 Jun. 2020.

[239] P.-L. Germain, A. Lun, C. G. Meixide, W. Macnair, and M. D. Robinson, "Doublet identification in single-cell sequencing data using scdblfinder," *F1000Research*, vol. 10, p. 979, May 2022.

[240] S. L. Wolock, R. Lopez, and A. M. Klein, "Scrublet: Computational identification of cell doublets in single-cell transcriptomic data," *Cell Systems*, vol. 8, 281–291.e9, 4 Apr. 2019.

[241] Y. Hao, S. Hao, E. Andersen-Nissen, *et al.*, "Integrated analysis of multimodal single-cell data," *Cell*, vol. 184, 3573–3587.e29, 13 Jun. 2021.

[242] T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, and R. Satija, "Single-cell chromatin state analysis with signac," *Nature Methods*, vol. 18, pp. 1333–1341, 11 Nov. 2021.

[243] D. Aran, A. P. Looney, L. Liu, *et al.*, "Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage," *Nature Immunology*, vol. 20, pp. 163–172, 2 Feb. 2019.

[244] D. A. Cusanovich, R. Daza, A. Adey, *et al.*, "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing," *Science*, vol. 348, pp. 910–914, 6237 May 2015.

[245] I. Korsunsky, N. Millard, J. Fan, *et al.*, "Fast, sensitive and accurate integration of single-cell data with harmony," *Nature Methods*, vol. 16, pp. 1289–1296, 12 Dec. 2019.

[246] N. A. Mabbott, J. Baillie, H. Brown, T. C. Freeman, and D. A. Hume, "An expression atlas of human primary cells: Inference of gene function from coexpression networks," *BMC Genomics*, vol. 14, p. 632, 1 2013.

[247] K. A. Oetjen, K. E. Lindblad, M. Goswami, *et al.*, "Human bone marrow assessment by single-cell rna sequencing, mass cytometry, and flow cytometry," *JCI Insight*, vol. 3, 23 Dec. 2018.

[248] J. M. Granja, S. Klemm, L. M. McGinnis, *et al.*, "Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia," *Nature Biotechnology*, vol. 37, pp. 1458–1465, 12 Dec. 2019.

[249] M. Li, X. Zhang, K. S. Ang, *et al.*, "Disco: A database of deeply integrated human single-cell omics data," *Nucleic Acids Research*, vol. 50, pp. D596–D602, D1 Jan. 2022.

[250] T. Stuart, A. Butler, P. Hoffman, *et al.*, "Comprehensive integration of single-cell data," *Cell*, vol. 177, 1888–1902.e21, 7 Jun. 2019.

[251] G. Wang, W. X. Wen, A. J. Mead, A. Roy, B. Psaila, and S. Thongjuea, "Processing single-cell rna-seq datasets using singcellar," *STAR Protocols*, vol. 3, p. 101 266, 2 Jun. 2022.

[252] S. Aibar, C. B. González-Blas, T. Moerman, *et al.*, "Scenic: Single-cell regulatory network inference and clustering," *Nature Methods*, vol. 14, pp. 1083–1086, 11 Nov. 2017.

[253] S. Z. Xie, K. B. Kaufmann, W. Wang, *et al.*, "Sphingosine-1-phosphate receptor 3 potentiates inflammatory programs in normal and leukemia stem cells to promote differentiation," *Blood Cancer Discovery*, vol. 2, pp. 32–53, 1 Jan. 2021.

[254] Y. Zhang, X. Xie, Y. Huang, *et al.*, "Temporal molecular program of human hematopoietic stem and progenitor cells after birth," *Developmental Cell*, vol. 57, 2745–2760.e6, 24 Dec. 2022.

[255] Y. Tanigawa, E. S. Dyer, and G. Bejerano, "Whichtf is functionally important in your open chromatin data?" *PLoS computational biology*, vol. 18, e1010378, 8 Aug. 2022.

[256] K. Street, D. Risso, R. B. Fletcher, *et al.*, "Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics," *BMC Genomics*, vol. 19, p. 477, 1 Dec. 2018.

[257] V. K. Kartha, F. M. Duarte, Y. Hu, *et al.*, "Functional inference of gene regulation using single-cell multi-omics," *Cell Genomics*, vol. 2, 9 Sep. 2022.

[258] R. Browaeys, W. Saelens, and Y. Saeys, "Nichenet: Modeling intercellular communication by linking ligands to target genes," *Nature Methods*, vol. 17, pp. 159–162, 2 Feb. 2020.

[259] R. Browaeys, J. Gilis, C. Sang-Aram, *et al.*, "Multinichenet: A flexible framework for differential cell-cell communication analysis from multi-sample multi-condition single-cell transcriptomics data," *bioRxiv*, p. 2023.06.13.544751, Jan. 2023. [Online]. Available: `http://biorxiv.org/content/early/2023/06/14/2023.06.13.544751.abstract`.

[260] J. Windhager, V. R. T. Zanotelli, D. Schulz, *et al.*, "An end-to-end workflow for multiplexed image processing and analysis," *Nature Protocols*, vol. 18, pp. 3565–3613, 11 Nov. 2023.

[261] S. Berg, D. Kutra, T. Kroeger, *et al.*, "Ilastik: Interactive machine learning for (bio)image analysis," *Nature Methods*, vol. 16, pp. 1226–1232, 12 Dec. 2019.

[262] P. Bankhead, M. B. Loughrey, J. A. Fernández, *et al.*, "Qupath: Open source software for digital pathology image analysis," *Scientific Reports*, vol. 7, p. 16 878, 1 Dec. 2017.

[263] Y. Feng, T. Yang, J. Zhu, *et al.*, "Spatial analysis with spiat and spasim to characterize and simulate tissue microenvironments," *Nature Communications*, vol. 14, p. 2697, 1 May 2023.

[264] K. B. Schoedel, M. N. F. Morcos, T. Zerjatke, *et al.*, "The bulk of the hematopoietic stem cell population is dispensable for murine steady-state and stress hematopoiesis," *Blood*, vol. 128, pp. 2285–2296, 19 Nov. 2016.

[265] A. E. Rodriguez-Fraticelli, S. L. Wolock, C. S. Weinreb, *et al.*, "Clonal analysis of lineage fate in native haematopoiesis," *Nature*, vol. 553, pp. 212–216, 7687 Jan. 2018.

[266] S. E. W. Jacobsen and C. Nerlov, "Haematopoiesis in the era of advanced single-cell technologies," *Nature Cell Biology*, vol. 21, pp. 2–8, 1 Jan. 2019.

[267] J. Carrelha, Y. Meng, L. M. Kettyle, *et al.*, "Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells," *Nature*, vol. 554, pp. 106–111, 7690 Feb. 2018.

[268] E. A. McCulloch and J. E. Till, "The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice," *Radiation Research*, vol. 13, p. 115, 1 Jul. 1960.

[269] D. Metcalf, "The colony stimulating factors discovery, development, and clinical applications," *Cancer*, vol. 65, pp. 2185–2195, 10 May 1990.

[270] G. J. Spangrude, S. Heimfeld, and I. L. Weissman, "Purification and characterization of mouse hematopoietic stem cells," *Science*, vol. 241, pp. 58–62, 4861 Jul. 1988.

[271] M. J. Kiel, Ö. H. Yilmaz, T. Iwashita, O. H. Yilmaz, C. Terhorst, and S. J. Morrison, "Slam family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells," *Cell*, vol. 121, pp. 1109–1121, 7 Jul. 2005.

[272] K. Akashi, D. Traver, T. Miyamoto, and I. L. Weissman, "A clonogenic common myeloid progenitor that gives rise to all myeloid lineages," *Nature*, vol. 404, pp. 193–197, 6774 Mar. 2000.

[273] T. Graf and T. Enver, "Forcing cells to change lineages," *Nature*, vol. 462, pp. 587–594, 7273 Dec. 2009.

[274] D. Metcalf, "Lineage commitment and maturation in hematopoietic cells: The case for extrinsic regulation," *Blood*, vol. 92, pp. 345–347, 2 Jul. 1998.

[275] C. E. Müller-Sieburg, R. H. Cho, M. Thoman, B. Adkins, and H. B. Sieburg, "Deterministic regulation of hematopoietic stem cell self-renewal and differentiation.," *Blood*, vol. 100, pp. 1302–9, 4 Aug. 2002.

[276] Y. Meng, J. Carrelha, R. Drissen, *et al.*, "Epigenetic programming defines haematopoietic stem cell fate restriction," *Nature Cell Biology*, vol. 25, pp. 812–822, 6 Jun. 2023.

[277] S. A. Lambert, A. Jolma, L. F. Campitelli, *et al.*, "The human transcription factors," *Cell*, vol. 172, pp. 650–665, 4 Feb. 2018.

[278] A. Visel, M. J. Blow, Z. Li, *et al.*, "Chip-seq accurately predicts tissue-specific activity of enhancers," *Nature*, vol. 457, pp. 854–858, 7231 Feb. 2009.

[279] C. Lancrin, P. Sroczynska, C. Stephenson, T. Allen, V. Kouskoff, and G. Lacaud, "The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage," *Nature*, vol. 457, pp. 892–895, 7231 Feb. 2009.

[280] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell*, vol. 126, pp. 663–676, 4 Aug. 2006.

[281] S. Ma, B. Zhang, L. M. LaFave, *et al.*, "Chromatin potential identified by shared single-cell profiling of rna and chromatin," *Cell*, vol. 183, 1103–1116.e20, 4 Nov. 2020.

[282] N. Gamble, A. Bradu, J. A. Caldwell, *et al.*, "Pu.1 and bcl11b sequentially cooperate with runx1 to anchor mswi/snf to poise the t cell effector landscape," *Nature Immunology*, vol. 25, pp. 860–872, 5 May 2024.

[283] M. K. Strasser, P. S. Hoppe, D. Loeffler, *et al.*, "Lineage marker synchrony in hematopoietic genealogies refutes the pu.1/gata1 toggle switch paradigm," *Nature Communications*, vol. 9, p. 2697, 1 Jul. 2018.

[284] E. M. Pietras, D. Reynaud, Y.-A. Kang, *et al.*, "Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions," *Cell Stem Cell*, vol. 17, pp. 35–46, 1 Jul. 2015.

[285] Y. Nagai, K. P. Garrett, S. Ohta, *et al.*, "Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment," *Immunity*, vol. 24, pp. 801–812, 6 Jun. 2006.

[286] K. Y. King and M. A. Goodell, "Inflammatory modulation of hscs: Viewing the hsc as a foundation for the immune response," *Nature Reviews Immunology*, vol. 11, pp. 685–692, 10 Oct. 2011.

[287] A. Illing, P. Liu, S. Ostermay, *et al.*, "Estradiol increases hematopoietic stem and progenitor cells independent of its actions on bone," *Haematologica*, vol. 97, pp. 1131–1135, 8 Aug. 2012.

[288] C. Morganti, N. Cabezas-Wallscheid, and K. Ito, "Metabolic regulation of hematopoietic stem cells," *HemaSphere*, vol. 6, e740, 7 Jul. 2022.

[289] Y.-A. Kang, H. Paik, S. Y. Zhang, *et al.*, "Secretory mpp3 reinforce myeloid differentiation trajectory and amplify myeloid cell production," *Journal of Experimental Medicine*, vol. 220, 8 Aug. 2023.

[290] A. Hérault, M. Binnewies, S. Leong, *et al.*, "Myeloid progenitor cluster formation drives emergency and leukaemic myelopoiesis," *Nature*, vol. 544, pp. 53–58, 7648 Apr. 2017.

[291] D. Reynaud, E. Pietras, K. Barry-Holson, *et al.*, "Il-6 controls leukemic multipotent progenitor cell fate and contributes to chronic myelogenous leukemia development," *Cancer Cell*, vol. 20, pp. 661–673, 5 Nov. 2011.

[292] J. W. Swann, L. A. Koneva, D. Regan-Komito, S. N. Sansom, F. Powrie, and T. Griseri, "Il-33 promotes anemia during chronic inflammation by inhibiting differentiation of erythroid progenitors," *Journal of Experimental Medicine*, vol. 217, 9 Sep. 2020.

[293] M. Raundhal, S. Ghosh, S. A. Myers, *et al.*, "Blockade of il-22 signaling reverses erythroid dysfunction in stress-induced anemias," *Nature Immunology*, vol. 22, pp. 520–529, 4 Apr. 2021.

[294] H. Kato, A. Itoh-Nakadai, M. Matsumoto, *et al.*, "Infection perturbs bach2- and bach1-dependent erythroid lineage 'choice' to cause anemia," *Nature Immunology*, vol. 19, pp. 1059–1070, 10 Oct. 2018.

[295] A. Itoh-Nakadai, M. Matsumoto, H. Kato, *et al.*, "A bach2-cebp gene regulatory network for the commitment of multipotent hematopoietic progenitors," *Cell Reports*, vol. 18, pp. 2401–2414, 10 Mar. 2017.

[296] H. Hirai, P. Zhang, T. Dayaram, *et al.*, "C/ebp is required for 'emergency' granulopoiesis," *Nature Immunology*, vol. 7, pp. 732–739, 7 Jul. 2006.

[297] L. C. Jones, M.-L. Lin, S.-S. Chen, *et al.*, "Expression of c/ebp from the c/ebp gene locus is sufficient for normal hematopoiesis in vivo," *Blood*, vol. 99, pp. 2032–2036, 6 Mar. 2002.

[298] B. de Laval, J. Maurizio, P. K. Kandalla, *et al.*, "C/ebp-dependent epigenetic memory induces trained immunity in hematopoietic stem cells," *Cell Stem Cell*, vol. 26, 657–674.e8, 5 2020. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S1934590920300175`.

[299] R. Zittoun, "Subacute and chronic myelomonocytic leukaemia: A distinct haematological entity," *British Journal of Haematology*, vol. 32, pp. 1–7, 1 Jan. 1976.

[300] T. Ruutu, S. Partanen, R. Lintula, L. Teerenhovi, and S. Knuutila, "Erythroid and granulocyte-macrophage colony formation in myelodysplastic syndromes," *Scandinavian Journal of Haematology*, vol. 32, pp. 395–402, 4 Apr. 1984.

[301] K. Geissler, E. Jäger, A. Barna, *et al.*, "Correlation of ras-pathway mutations and spontaneous myeloid colony growth with progression and transformation in chronic myelomonocytic leukemia—a retrospective analysis in 337 patients," *International Journal of Molecular Sciences*, vol. 21, p. 3025, 8 Apr. 2020.

[302] A. Beke, L. Laplane, J. Riviere, *et al.*, "Multilayer intraclonal heterogeneity in chronic myelomonocytic leukemia," *Haematologica*, vol. 105, pp. 112–123, 1 Jan. 2020.

[303] S. Yamazaki, K. Taoka, S. Arai, *et al.*, "Patient-derived induced pluripotent stem cells identified slitrk4 as a causative gene of chronic myelomonocytic leukemia," *Blood*, vol. 128, pp. 1134–1134, 22 Dec. 2016.

[304] K. Taoka, S. Arai, K. Kataoka, *et al.*, "Using patient-derived ipscs to develop humanized mouse models for chronic myelomonocytic leukemia and therapeutic drug identification, including liposomal clodronate," *Scientific Reports*, vol. 8, p. 15 855, 1 Oct. 2018.

[305] D. H. Wiseman, S. M. Baker, A. V. Dongre, *et al.*, "Chronic myelomonocytic leukaemia stem cell transcriptomes anticipate disease morphology and outcome," *EBioMedicine*, vol. 58, Aug. 2020.

[306] K. Gurashi, J. Weightman, S. M. Baker, *et al.*, "Multiomic single-cell sequencing reveals patterns of disease evolution and acute transformation in chronic myelomonocytic leukaemia," *Blood*, vol. 138, pp. 2586–2586, Supplement 1 Nov. 2021.

[307] A. Zehir, R. Benayed, R. H. Shah, *et al.*, "Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.," *Nature medicine*, vol. 23, pp. 703–713, 6 Jun. 2017.

[308] D. A. Arber, A. Orazi, R. Hasserjian, *et al.*, "The 2016 revision to the world health organization classification of myeloid neoplasms and acute leukemia," *Blood*, vol. 127, pp. 2391–2405, 20 May 2016.

[309] M. A. Gillespie, C. G. Palii, D. Sanchez-Taltavull, *et al.*, "Absolute quantification of transcription factors reveals principles of gene regulation in erythropoiesis," *Molecular Cell*, vol. 78, 960–974.e11, 5 Jun. 2020.

[310] M. Karin, Z.-g. Liu, and E. Zandi, "Ap-1 function and regulation," *Current Opinion in Cell Biology*, vol. 9, pp. 240–246, 2 Apr. 1997.

[311] S. Okada, T. Fukuda, K. Inada, and T. Tokuhisa, "Prolonged expression of c-fos suppresses cell cycle entry of dormant hematopoietic stem cells," *Blood*, vol. 93, pp. 816–825, 3 Feb. 1999.

[312] M. Santaguida, K. Schepers, B. King, *et al.*, "Junb protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal," *Cancer Cell*, vol. 15, pp. 341–352, 4 Apr. 2009.

[313] L. Velten, S. F. Haas, S. Raffel, *et al.*, "Human haematopoietic stem cell lineage commitment is a continuous process," *Nature Cell Biology*, vol. 19, pp. 271–281, 4 Mar. 2017.

[314] D.-M. Popescu, R. A. Botting, E. Stephenson, *et al.*, "Decoding human fetal liver haematopoiesis.," *Nature*, vol. 574, pp. 365–371, 7778 Oct. 2019.

[315] R. Drissen, S. Thongjuea, K. Theilgaard-Mönch, and C. Nerlov, "Identification of two distinct pathways of human myelopoiesis.," *Science immunology*, vol. 4, 35 May 2019.

[316] A. Itoh-Nakadai, R. Hikota, A. Muto, *et al.*, "The transcription repressors bach2 and bach1 promote b cell development by repressing the myeloid program," *Nature Immunology*, vol. 15, pp. 1171–1180, 12 Dec. 2014.

[317] G. Montalban-Bravo, N. Thongon, J. J. Rodriguez-Sevilla, *et al.*, "Targeting mcl1-driven anti-apoptotic pathways overcomes blast progression after hypomethylating agent failure in chronic myelomonocytic leukemia," *Cell Reports Medicine*, vol. 5, p. 101 585, 6 Jun. 2024.

[318] Q. Shen, J. Ouyang, G. Tang, *et al.*, "Flow cytometry immunophenotypic findings in chronic myelomonocytic leukemia and its utility in monitoring treatment response," *European Journal of Haematology*, vol. 95, pp. 168–176, 2 Aug. 2015.

[319] F. Lyko, "The dna methyltransferase family: A versatile toolkit for epigenetic regulation," *Nature Reviews Genetics*, vol. 19, pp. 81–92, 2 Feb. 2018.

[320] S. K. Sayeed, J. Zhao, B. K. Sathyanarayana, J. P. Golla, and C. Vinson, "C/ebp (cebpb) protein binding to the c/ebp|cre dna 8-mer ttgc|gtca is inhibited by 5hmc and enhanced by 5mc, 5fc, and 5cac in the cg dinucleotide," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1849, pp. 583–589, 6 Jun. 2015.

[321] M. d Mesbah Uddin, N. Q. H. Nguyen, B. Yu, *et al.*, "Clonal hematopoiesis of indeterminate potential, dna methylation, and risk for coronary artery disease," *Nature Communications*, vol. 13, p. 5350, 1 Sep. 2022.

[322] J. W. Swann, O. C. Olson, and E. Passegué, "Made to order: Emergency myelopoiesis and demand-adapted innate immune cell production," *Nature Reviews Immunology*, vol. 24, pp. 596–613, 8 Aug. 2024.

[323] C. FRANCESCHI, M. BONAFÈ, S. VALENSIN, *et al.*, "Inflamm-aging: An evolutionary perspective on immunosenescence," *Annals of the New York Academy of Sciences*, vol. 908, pp. 244–254, 1 Jun. 2000.

[324] W. Kuribayashi, M. Oshima, N. Itokawa, *et al.*, "Limited rejuvenation of aged hematopoietic stem cells in young bone marrow niche," *Journal of Experimental Medicine*, vol. 218, 3 Mar. 2021.

[325] J. L. Rabe, G. Hernandez, J. S. Chavez, T. S. Mills, C. Nerlov, and E. M. Pietras, "Cd34 and epcr coordinately enrich functional murine hematopoietic stem cells under normal and inflammatory conditions," *Experimental Hematology*, vol. 81, 1–15.e6, Jan. 2020.

[326] H. K. A. Mikkola and S. H. Orkin, "The journey of developing hematopoietic stem cells," *Development*, vol. 133, pp. 3733–3744, 19 Oct. 2006.

[327] H. J. Sutherland, C. J. Eaves, P. M. Lansdorp, J. D. Thacker, and D. E. Hogge, "Differential regulation of primitive human hematopoietic cells in long-term cultures maintained on genetically engineered murine stromal cells.," *Blood*, vol. 78, pp. 666–72, 3 Aug. 1991.

[328] A. J. Naiyer, D. Y. Jo, J. Ahn, *et al.*, "Stromal derived factor-1-induced chemokinesis of cord blood cd34(+) cells (long-term culture-initiating cells) through endothelial cells is mediated by e-selectin.," *Blood*, vol. 94, pp. 4011–9, 12 Dec. 1999.

[329] N. Nishi, R. Ishikawa, H. Inoue, *et al.*, "Granulocyte-colony stimulating factor and stem cell factor are the crucial factors in long-term culture of human primitive hematopoietic cells supported by a murine stromal cell line.," *Experimental hematology*, vol. 24, pp. 1312–21, 11 Sep. 1996.

[330] T. Reya, A. W. Duncan, L. Ailles, *et al.*, "A role for wnt signalling in self-renewal of haematopoietic stem cells," *Nature*, vol. 423, pp. 409–414, 6938 May 2003.

[331] T. E. North, W. Goessling, C. R. Walkley, *et al.*, "Prostaglandin e2 regulates vertebrate haematopoietic stem cell homeostasis," *Nature*, vol. 447, pp. 1007–1011, 7147 Jun. 2007.

[332] A. C. Wilkinson, R. Ishida, M. Kikuchi, *et al.*, "Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation," *Nature*, vol. 571, pp. 117–121, 7763 Jul. 2019.

[333] S. Bandyopadhyay, M. P. Duffy, K. J. Ahn, *et al.*, "Mapping the cellular biogeography of human bone marrow niches using single-cell transcriptomics and proteomic imaging," *Cell*, vol. 187, 3120–3140.e29, 12 Jun. 2024.

[334] I. G. Winkler, V. Barbier, R. Wadley, A. C. W. Zannettino, S. Williams, and J.-P. Lévesque, "Positioning of bone marrow hematopoietic and stromal cells relative to blood flow in vivo: Serially reconstituting hematopoietic stem cells reside in distinct nonperfused niches," *Blood*, vol. 116, pp. 375–385, 3 Jul. 2010.

[335] S. Lymperi, F. Ferraro, and D. T. Scadden, "The hsc niche concept has turned 31," *Annals of the New York Academy of Sciences*, vol. 1192, pp. 12–18, 1 Apr. 2010.

[336] J. Zhang, Q. Wu, C. B. Johnson, *et al.*, "In situ mapping identifies distinct vascular niches for myelopoiesis," *Nature*, vol. 590, pp. 457–462, 7846 Feb. 2021.

[337] Q. Wu, J. Zhang, S. Kumar, *et al.*, "Resilient anatomy and local plasticity of naive and stress haematopoiesis," *Nature*, vol. 627, pp. 839–846, 8005 Mar. 2024.

[338] M. B. Buechler, H. M. Akilesh, and J. A. Hamerman, "Cutting edge: Direct sensing of tlr7 ligands and type i ifn by the common myeloid progenitor promotes mtor/pi3k-dependent emergency myelopoiesis," *The Journal of Immunology*, vol. 197, pp. 2577–2582, 7 Oct. 2016.

[339] P. K. Kandalla, S. Sarrazin, K. Molawi, *et al.*, "M-csf improves protection against bacterial and fungal infections after hematopoietic stem/progenitor cell transplantation," *Journal of Experimental Medicine*, vol. 213, pp. 2269–2279, 11 Oct. 2016.

[340] I. Mitroulis, K. Ruppova, B. Wang, *et al.*, "Modulation of myelopoiesis progenitors is an integral component of trained immunity," *Cell*, vol. 172, 147–161.e12, 1-2 Jan. 2018.

[341] P. R. Nagareddy, M. Kraakman, S. L. Masters, *et al.*, "Adipose tissue macrophages promote myelopoiesis and monocytosis in obesity," *Cell Metabolism*, vol. 19, pp. 821–835, 5 May 2014.

[342] P. Valent, A. Orazi, M. R. Savona, *et al.*, "Proposed diagnostic criteria for classical chronic myelomonocytic leukemia (cmml), cmml variants and pre-cmml conditions," *Haematologica*, vol. 104, pp. 1935–1949, 10 Oct. 2019.

[343] A. Orazi, R. Chiu, D. P. O'Malley, *et al.*, "Chronic myelomonocytic leukemia: The role of bone marrow biopsy immunohistology," *Modern Pathology*, vol. 19, pp. 1536–1545, 12 Jan. 2006.

[344] H.-P. Horny, K. Sotlar, W. R. Sperr, and P. Valent, "Systemic mastocytosis with associated clonal haematological non-mast cell lineage diseases: A histopathological challenge," *Journal of Clinical Pathology*, vol. 57, pp. 604–608, 6 Jun. 2004.

[345] R. Xu, X. Huang, C. Li, *et al.*, "Bone marrow mesenchymal stromal cells in chronic myelomonocytic leukaemia: Overactivated wnt/-catenin signalling by parallel rna sequencing and dysfunctional phenotypes," *British Journal of Haematology*, vol. 193, pp. 928–940, 5 Jun. 2021.

[346] D. K. Edwards, K. Watanabe-Smith, A. Rofelty, *et al.*, "Csf1r inhibitors exhibit antitumor activity in acute myeloid leukemia by blocking paracrine signals from support cells," *Blood*, vol. 133, pp. 588–599, 6 Feb. 2019.

[347] I. E. Stergiou, K. Kambas, A. Poulaki, *et al.*, "Exploiting the role of hypoxia-inducible factor 1 and pseudohypoxia in the myelodysplastic syndrome pathophysiology.," *International journal of molecular sciences*, vol. 22, 8 Apr. 2021.

[348] H. Shi, Y. Wang, R. Li, *et al.*, "Alteration in the cytokine secretion of bone marrow stromal cells from patients with chronic myelomonocytic leukemia contribute to impaired hematopoietic supportive activity," *Stem Cells International*, vol. 2018, pp. 1–9, Jul. 2018.

[349] Z. Poon, N. Dighe, S. S. Venkatesan, *et al.*, "Bone marrow mscs in mds: Contribution towards dysfunctional hematopoiesis and potential targets for disease response to hypomethylating therapy," *Leukemia*, vol. 33, pp. 1487–1500, 6 Jun. 2019.

[350] L. Dong, W.-M. Yu, H. Zheng, *et al.*, "Leukaemogenic effects of ptpn11 activating mutations in the stem cell microenvironment.," *Nature*, vol. 539, pp. 304–308, 7628 Nov. 2016.

[351] K. Geissler, W. Hinterberger, P. Bettelheim, O. Haas, and K. Lechner, "Colony growth characteristics in chronic myelomonocytic leukemia," *Leukemia Research*, vol. 12, pp. 373–377, 5 Jan. 1988.

[352] M. Everson, C. Brown, and M. Lilly, "Interleukin-6 and granulocyte-macrophage colony-stimulating factor are candidate growth factors for chronic myelomonocytic leukemia cells," *Blood*, vol. 74, pp. 1472–1476, 5 Oct. 1989.

[353] L. Oehler, M. Foedinger, M. Koeller, *et al.*, "Interleukin-10 inhibits spontaneous colony-forming unit–granulocyte-macrophage growth from human peripheral blood mononuclear cells by suppression of endogenous granulocyte-macrophage colony-stimulating factor release," *Blood*, vol. 89, pp. 1147–1153, 4 Feb. 1997.

[354] C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, and A. M. Klein, "Lineage tracing on transcriptional landscapes links state to fate during differentiation," *Science*, vol. 367, 6479 Feb. 2020.

[355] A. Mildner, K.-W. Kim, and S. Yona, "Unravelling monocyte functions: From the guardians of health to the regulators of disease," *Discovery Immunology*, vol. 3, 1 Jan. 2024.

[356] W. Li, Y. Wang, H. Zhao, *et al.*, "Identification and transcriptome analysis of erythroblastic island macrophages," *Blood*, vol. 134, pp. 480–491, 5 Aug. 2019.

[357] Y. Luo, L. Shao, J. Chang, *et al.*, "M1 and m2 macrophages differentially regulate hematopoietic stem cell self-renewal and ex vivo expansion," *Blood Advances*, vol. 2, pp. 859–870, 8 Apr. 2018.

[358] C. P. Rodrigues, J. M. Collins, S. Yang, *et al.*, "Transcripts of repetitive dna elements signal to block phagocytosis of hematopoietic stem cells," *Science*, vol. 385, 6714 Sep. 2024.

[359] L. P. D'Atri, R. G. Pozner, K. A. Nahmod, *et al.*, "Paracrine regulation of megakaryo/thrombopoiesis by macrophages," *Experimental Hematology*, vol. 39, pp. 763–772, 7 Jul. 2011.

[360] D. Selimoglu-Buet, O. Wagner-Ballon, V. Saada, *et al.*, "Characteristic repartition of monocyte subsets as a diagnostic signature of chronic myelomonocytic leukemia," *Blood*, vol. 125, pp. 3618–3626, 23 Jun. 2015.

[361] C. A. Hudson, W. R. Burack, P. C. Leary, and J. M. Bennett, "Clinical utility of classical and nonclassical monocyte percentage in the diagnosis of chronic myelomonocytic leukemia," *American Journal of Clinical Pathology*, vol. 150, pp. 293–302, 4 Aug. 2018.

[362] S. Tarfi, V. Harrivel, F. Dumezy, *et al.*, "Multicenter validation of the flow measurement of classical monocyte fraction for chronic myelomonocytic leukemia diagnosis," *Blood Cancer Journal*, vol. 8, p. 114, 11 Nov. 2018.

[363] L. Barge, M. Gooch, M. Hendle, and E. Simleit, "Real world implementation of flow cytometric monocyte subset partitioning for distinguishing chronic myelomonocytic leukaemia from other causes of monocytosis," *Pathology*, vol. 55, pp. 827–834, 6 Oct. 2023.

[364] X. Calvo, I. Parraga, N. Garcia-Gisbert, *et al.*, "High diagnostic utility of flow cytometry based peripheral blood monocyte subset analysis, cd56 and cd2 expression in chronic myelomonocytic leukemia (cmml)," *Blood*, vol. 134, pp. 5437–5437, Supplement$_1$ Nov. 2019.

[365] K.-m. Lee, J. M. Giltnane, J. M. Balko, *et al.*, "Myc and mcl1 cooperatively promote chemotherapy-resistant breast cancer stem cells via regulation of mitochondrial oxidative phosphorylation," *Cell Metabolism*, vol. 26, 633–647.e7, 4 Oct. 2017.

[366] H. M. Addinsell, R. Cant, N. J. Hull, *et al.*, "Multi-omic analysis of chronic myelomonocytic leukemia monocytes reveals metabolic and immune dysregulation leading to altered macrophage polarization," *Leukemia*, vol. 39, pp. 770–774, 3 Mar. 2025.

[367] A. Viola, F. Munari, R. Sánchez-Rodríguez, T. Scolaro, and A. Castegna, "The metabolic signature of macrophage responses," *Frontiers in Immunology*, vol. 10, Jul. 2019.

[368] N. J. Hull, R. Cant, L. A. Guest, *et al.*, "Granulocyte derived resistin inhibits monocyte maturation and induces immune suppression in cmml," Mar. 2025.

[369] A. Kentsis, C. Reed, K. L. Rice, *et al.*, "Autocrine activation of the met receptor tyrosine kinase in acute myeloid leukemia," *Nature Medicine*, vol. 18, pp. 1118–1122, 7 Jul. 2012.

[370] O. Bruserud, B. T. Gjertsen, O. T. Brustugun, *et al.*, "Effects of interleukin 10 on blast cells derived from patients with acute myelogenous leukemia.," *Leukemia*, vol. 9, pp. 1910–20, 11 Nov. 1995.

[371] M. G. Netea, C. A. Nold-Petry, M. F. Nold, *et al.*, "Differential requirement for the activation of the inflammasome for processing and release of il-1 in monocytes and macrophages," *Blood*, vol. 113, pp. 2324–2335, 10 Mar. 2009.

[372] M. Mussbacher, M. Derler, J. Basílio, and J. A. Schmid, "Nf-b in monocytes and macrophages – an inflammatory master regulator in multitalented immune cells," *Frontiers in Immunology*, vol. 14, Feb. 2023.

[373] P. Tuijnenburg, H. L. Allen, S. O. Burns, *et al.*, "Loss-of-function nuclear factor b subunit 1 (nfkb1) variants are the most common monogenic cause of common variable immunodeficiency in europeans," *Journal of Allergy and Clinical Immunology*, vol. 142, pp. 1285–1296, 4 Oct. 2018.

[374] D. Somma, F. O. Kok, D. Kerrigan, C. A. Wells, and R. J. Carmody, "Defining the role of nuclear factor (nf)-b p105 subunit in human macrophage by transcriptomic analysis of nfkb1 knockout thp1 cells," *Frontiers in Immunology*, vol. 12, Oct. 2021.

[375] P. Sojitra, P. Gandhi, P. Fitting, *et al.*, "Chronic myelomonocytic leukemia monocytes uniformly display a population of monocytes with cd11c underexpression," *American Journal of Clinical Pathology*, vol. 140, pp. 686–692, 5 Nov. 2013.

[376] D. Selimoglu-Buet, B. Badaoui, E. Benayoun, *et al.*, "Accumulation of classical monocytes defines a subgroup of mds that frequently evolves into cmml," *Blood*, vol. 130, pp. 832–835, 6 Aug. 2017.

[377] C.-F. Yang, J.-P. Gau, L.-T. Hsiao, and C.-Y. Hsu, "Clinical significance of blast percentage assessed by bone marrow trephine biopsy and aspirate smear of myeloid malignancies," *Pathology*, vol. 53, pp. 740–745, 6 Oct. 2021.

[378] V. F. I. Sangiorgio, D. A. Arber, and A. Orazi, "How i investigate chronic myelomonocytic leukemia," *International Journal of Laboratory Hematology*, vol. 42, pp. 101–108, 2 Apr. 2020.

[379] E. Sitnicka, F. W. Ruscetti, G. V. Priestley, N. S. Wolf, and S. H. Bartelmez, "Transforming growth factor beta 1 directly and reversibly inhibits the initial cell divisions of long-term repopulating hematopoietic stem cells.," *Blood*, vol. 88, pp. 82–8, 1 Jul. 1996.

[380] J. J. Letterio, A. G. Geiser, A. B. Kulkarni, *et al.*, "Autoimmunity associated with tgf-beta1-deficiency in mice is dependent on mhc class ii antigen expression.," *Journal of Clinical Investigation*, vol. 98, pp. 2109–2119, 9 Nov. 1996.

[381] K. Mitchell, L. Barreyro, T. I. Todorova, *et al.*, "Il1rap potentiates multiple oncogenic signaling pathways in aml," *Journal of Experimental Medicine*, vol. 215, pp. 1709–1727, 6 Jun. 2018.

[382] I. Ganan-Gomez, K. S. Chien, F. Ma, *et al.*, "The transcriptional and epigenetic reprogramming of aged hematopoietic stem cells drives myeloid rewiring in clonal hematopoiesis-associated cytopenias," *Blood*, vol. 138, pp. 3273–3273, Supplement 1 Nov. 2021.

[383] G. Garcia-Manero, V. Adema, S. Urrutia, *et al.*, "Clinical and biological effects of canakinumab in lower-risk myelodysplastic syndromes (mds): Results from a phase 2 clinical trial," *Blood*, vol. 140, pp. 2078–2080, Supplement 1 Nov. 2022.

[384] C. B. Marsh, R. P. Pomerantz, J. M. Parker, *et al.*, "Regulation of monocyte survival in vitro by deposited igg: Role of macrophage colony-stimulating factor," *The Journal of Immunology*, vol. 162, pp. 6217–6225, 10 May 1999.

[385] A. Tamura, H. Hirai, A. Yokota, *et al.*, "C/ebp is required for survival of ly6c(-) monocytes," *Blood*, vol. 130, pp. 1809–1818, 16 2017.

[386] Y. Xu, R. W. McKenna, N. J. Karandikar, A. J. Pildain, and S. H. Kroft, "Flow cytometric analysis of monocytes as a tool for distinguishing chronic myelomonocytic leukemia from reactive monocytosis," *American Journal of Clinical Pathology*, vol. 124, pp. 799–806, 5 Nov. 2005.

[387] X. Wang, H. Hisha, S. Taketani, *et al.*, "Neural cell adhesion molecule contributes to hemopoiesis-supporting capacity of stromal cell lines," *STEM CELLS*, vol. 23, pp. 1389–1399, 9 Oct. 2005.

[388] P. S. Walmod, K. Kolkova, V. Berezin, and E. Bock, "Zippers make signals: Ncam-mediated molecular interactions and signal transduction.," *Neurochemical research*, vol. 29, pp. 2015–35, 11 Nov. 2004.

[389] P. Milev, P. Maurel, M. Häring, R. K. Margolis, and R. U. Margolis, "Tag-1/axonin-1 is a high-affinity ligand of neurocan, phosphacan/protein-tyrosine phosphatase-zeta/beta, and n-cam.," *The Journal of biological chemistry*, vol. 271, pp. 15 716–23, 26 Jun. 1996.

[390] G. Paratcha, F. Ledda, and C. F. Ibáñez, "The neural cell adhesion molecule ncam is an alternative signaling receptor for gdnf family ligands.," *Cell*, vol. 113, pp. 867–79, 7 Jun. 2003.

[391] C. Francavilla, P. Cattaneo, V. Berezin, *et al.*, "The binding of ncam to fgfr1 induces a specific cellular response mediated by receptor trafficking," *Journal of Cell Biology*, vol. 187, pp. 1101–1116, 7 Dec. 2009.

[392] S. D. Storms, A. C. Kim, B. H. Tran, G. J. Cole, and B. A. Murray, "Ncam-mediated adhesion of transfected cells to agrin.," *Cell adhesion and communication*, vol. 3, pp. 497–509, 6 Apr. 1996.

[393] V. Dardalhon, A. C. Anderson, J. Karman, *et al.*, "Tim-3/galectin-9 pathway: Regulation of th1 immunity through promotion of cd11b+ly-6g+ myeloid cells," *The Journal of Immunology*, vol. 185, pp. 1383–1392, 3 Aug. 2010.

[394] M. A. Dao, K. Hashino, I. Kato, and J. A. Nolta, "Adhesion to fibronectin maintains regenerative capacity during ex vivo culture and transduction of human hematopoietic stem and progenitor cells.," *Blood*, vol. 92, pp. 4612–21, 12 Dec. 1998.

[395] T. Matsunaga, N. Takemoto, T. Sato, *et al.*, "Interaction between leukemic-cell vla-4 and stromal fibronectin is a decisive factor for minimal residual disease of acute myelogenous leukemia," *Nature Medicine*, vol. 9, pp. 1158–1165, 9 Sep. 2003.

[396] F. Galimi, G. P. Bagnara, L. Bonsi, *et al.*, "Hepatocyte growth factor induces proliferation and differentiation of multipotent and erythroid hemopoietic progenitors.," *The Journal of cell biology*, vol. 127, pp. 1743–1754, 6 Dec. 1994.

[397] T. E. Kmiecik, J. R. Keller, E. Rosen, and G. F. V. Woude, "Hepatocyte growth factor is a synergistic factor for the growth of hematopoietic progenitor cells.," *Blood*, vol. 80, pp. 2454–7, 10 Nov. 1992.

[398] S. Neuss, E. Becher, M. Wöltje, L. Tietze, and W. Jahnen-Dechent, "Functional expression of hgf and hgf receptor/c-met in adult human mesenchymal stem cells suggests a role in cell mobilization, tissue repair, and wound healing," *STEM CELLS*, vol. 22, pp. 405–414, 3 May 2004.

[399] E. C. Chen, H. Gandler, I. Tošić, *et al.*, "Targeting met and fgfr in relapsed or refractory acute myeloid leukemia: Preclinical and clinical findings, and signal transduction correlates," *Clinical Cancer Research*, vol. 29, pp. 878–887, 5 Mar. 2023.

[400] E. Armingol, A. Officer, O. Harismendy, and N. E. Lewis, "Deciphering cell–cell interactions and communication from gene expression," *Nature Reviews Genetics*, vol. 22, pp. 71–88, 2 Feb. 2021.

[401] J. P. Annes, J. S. Munger, and D. B. Rifkin, "Making sense of latent tgf activation," *Journal of Cell Science*, vol. 116, pp. 217–224, 2 Jan. 2003.

[402] R. O. Hynes, "The extracellular matrix: Not just pretty fibrils," *Science*, vol. 326, pp. 1216–1219, 5957 Nov. 2009.

[403] Z. Deng, T. Fan, C. Xiao, *et al.*, "Tgf- signaling in health, disease and therapeutics," *Signal Transduction and Targeted Therapy*, vol. 9, p. 61, 1 Mar. 2024.

[404] J. J. Rodriguez-Sevilla, V. Adema, K. S. Chien, *et al.*, "The il-1 inhibitor canakinumab in previously treated lower-risk myelodysplastic syndromes: A phase 2 clinical trial," *Nature Communications*, vol. 15, p. 9840, 1 Nov. 2024.

[405] T. Bradley and D. Metcalf, "The growth of mouse bone marrow cells *IN VITRO*," *Australian Journal of Experimental Biology and Medical Science*, vol. 44, pp. 287–300, 3 Jun. 1966.

[406] E. O'Reilly, H. A. Zeinabad, and E. Szegezdi, "Hematopoietic versus leukemic stem cell quiescence: Challenges and therapeutic opportunities," *Blood Reviews*, vol. 50, p. 100 850, Nov. 2021.

[407] M. Endele, M. Etzrodt, and T. Schroeder, "Instruction of hematopoietic lineage choice by cytokine signaling," *Experimental Cell Research*, vol. 329, pp. 207–213, 2 Dec. 2014.

[408] E. R. Stanley and P. M. Heard, "Factors regulating macrophage production and growth. purification and some properties of the colony stimulating factor from medium conditioned by mouse l cells.," *The Journal of biological chemistry*, vol. 252, pp. 4305–12, 12 Jun. 1977.

[409] E. R. Stanley, M. Cifone, P. M. Heard, and V. Defendi, "Factors regulating macrophage production and growth: Identity of colony-stimulating factor and macrophage growth factor.," *The Journal of experimental medicine*, vol. 143, pp. 631–647, 3 Mar. 1976.

[410] X.-M. Dai, X.-H. Zong, V. Sylvestre, and E. R. Stanley, "Incomplete restoration of colony-stimulating factor 1 (csf-1) function in csf-1–deficient csf1op/csf1op mice by transgenic expression of cell surface csf-1," *Blood*, vol. 103, pp. 1114–1123, 3 Feb. 2004.

[411] C. J. Sherr, C. W. Rettenmier, R. Sacca, M. F. Roussel, A. T. Look, and E. R. Stanley, "The c-fms proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, csf-1.," *Cell*, vol. 41, pp. 665–76, 3 Jul. 1985.

[412] H. Lin, E. Lee, K. Hestir, *et al.*, "Discovery of a cytokine and its receptor by functional screening of the extracellular proteome.," *Science (New York, N.Y.)*, vol. 320, pp. 807–11, 5877 May 2008.

[413] P. S. Lee, Y. Wang, M. G. Dominguez, *et al.*, "The cbl protooncoprotein stimulates csf-1 receptor multiubiquitination and endocytosis, and attenuates macrophage proliferation.," *The EMBO journal*, vol. 18, pp. 3616–28, 13 Jul. 1999.

[414] E. R. Stanley and V. Chitu, "Csf-1 receptor signaling in myeloid cells.," *Cold Spring Harbor perspectives in biology*, vol. 6, 6 Jun. 2014.

[415] L. Bencheikh, M. K. Diop, J. Rivière, *et al.*, "Dynamic gene regulation by nuclear colony-stimulating factor 1 receptor in human monocytes and macrophages," *Nature Communications*, vol. 10, 1 Dec. 2019.

[416] N. Mossadegh-Keller, S. Sarrazin, P. K. Kandalla, *et al.*, "M-csf instructs myeloid lineage fate in single haematopoietic stem cells," *Nature*, vol. 497, pp. 239–243, 7448 2013.

[417] H. Krysinska, M. Hoogenkamp, R. Ingram, *et al.*, "A two-step, pu.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the c- *fms* gene," *Molecular and Cellular Biology*, vol. 27, pp. 878–887, 3 Feb. 2007.

[418] A. J. Fleetwood, T. Lawrence, J. A. Hamilton, and A. D. Cook, "Granulocyte-macrophage colony-stimulating factor (csf) and macrophage csf-dependent macrophage phenotypes display differences in cytokine profiles and transcription factor activities: Implications for csf blockade in inflammation," *The Journal of Immunology*, vol. 178, pp. 5245–5252, 8 Apr. 2007.

[419] K. P. A. MacDonald, V. Rowe, H. M. Bofinger, *et al.*, "The colony-stimulating factor 1 receptor is expressed on dendritic cells during differentiation and regulates their expansion," *The Journal of Immunology*, vol. 175, pp. 1399–1405, 3 Aug. 2005.

[420] S. Nandi, S. Gokhan, X.-M. Dai, *et al.*, "The csf-1 receptor ligands il-34 and csf-1 exhibit distinct developmental brain expression patterns and regulate neural progenitor cell maintenance and maturation," *Developmental Biology*, vol. 367, pp. 100–113, 2 Jul. 2012.

[421] J. W. Pollard and E. R. Stanley, "Pleiotropic roles for csf-1 in development defined by the mouse mutation osteopetrotic," in 1996, pp. 153–193.

[422] D. Huynh, X.-M. Dai, S. Nandi, *et al.*, "Colony stimulating factor-1 dependence of paneth cell development in the mouse small intestine," *Gastroenterology*, vol. 137, 136–144.e3, 1 Jul. 2009.

[423] X.-M. Dai, G. R. Ryan, A. J. Hapel, *et al.*, "Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects," *Blood*, vol. 99, pp. 111–120, 1 Jan. 2002.

[424] R. Rademakers, M. Baker, A. M. Nicholson, *et al.*, "Mutations in the colony stimulating factor 1 receptor (csf1r) gene cause hereditary diffuse leukoencephalopathy with spheroids.," *Nature genetics*, vol. 44, pp. 200–5, 2 Dec. 2011.

[425] C. Tomassetti, G. Insinga, F. Gimigliano, A. Morrione, A. Giordano, and E. Giurisato, "Insights into csf-1r expression in the tumor microenvironment.," *Biomedicines*, vol. 12, 10 Oct. 2024.

[426] M. Ciomei, L. Zanetta, F. Lussana, *et al.*, "A phase i/ii study of nms-03592088, a flt3, kit and csf1r inhibitor, in patients with relapsed or refractory aml or cmml," *Blood*, vol. 134, p. 3855, Supplement$_1$ 2019. [Online]. Available: `https://doi.org/10.1182/blood-2019-121954`.

[427] A. Curti, A. Rambaldi, C. Cattaneo, *et al.*, "Abstract ct025: Nms-03592088, a novel, potent flt3, kit and csf1r inhibitor with activity in flt3 positive acute myeloid leukemia patients with prior flt3 inhibitor experience," *Cancer Research*, vol. 83, CT025–CT025, 8$_S$*upplement* Apr. 2023.

[428] J. W. Tyner, C. E. Tognon, D. Bottomly, *et al.*, "Functional genomic landscape of acute myeloid leukaemia.," *Nature*, vol. 562, pp. 526–531, 7728 Oct. 2018.

[429] J. M. Ellegast, P. J. Rauch, L. V. Kovtonyuk, *et al.*, "Inv(16) and npm1mut amls engraft human cytokine knock-in mice," *Blood*, vol. 128, pp. 2130–2134, 17 Oct. 2016.

[430] T.-l. Gu, T. Mercher, J. W. Tyner, *et al.*, "A novel fusion of rbm6 to csf1r in acute megakaryoblastic leukemia.," *Blood*, vol. 110, pp. 323–33, 1 Jul. 2007.

[431] C. H. Ries, M. A. Cannarile, S. Hoves, *et al.*, "Targeting tumor-associated macrophages with anti-csf-1r antibody reveals a strategy for cancer therapy," *Cancer Cell*, vol. 25, pp. 846–859, 6 Jun. 2014.

[432] P. A. Cassier, A. Italiano, C. Gomez-Roca, *et al.*, "Long-term clinical activity, safety and patient-reported quality of life for emactuzumab-treated patients with diffuse-type tenosynovial giant-cell tumour," *European Journal of Cancer*, vol. 141, pp. 162–170, Dec. 2020.

[433] J. Massagué, "Tgf- signal transduction," *Annual Review of Biochemistry*, vol. 67, pp. 753–791, 1 Jun. 1998.

[434] S. Yamazaki, A. Iwama, S.-i. Takayanagi, K. Eto, H. Ema, and H. Nakauchi, "Tgf- as a candidate bone marrow niche signal to induce hematopoietic stem cell hibernation," *Blood*, vol. 113, pp. 1250–1256, 6 Feb. 2009.

[435] G. A. Challen, N. C. Boles, S. M. Chambers, and M. A. Goodell, "Distinct hematopoietic stem cell subtypes are differentially regulated by tgf-1," *Cell Stem Cell*, vol. 6, pp. 265–278, 3 Mar. 2010.

[436] Y. Zermati, S. Fichelson, F. Valensi, *et al.*, "Transforming growth factor inhibits erythropoiesis by blocking proliferation and accelerating differentiation of erythroid progenitors," *Experimental Hematology*, vol. 28, pp. 885–894, 8 Aug. 2000.

[437] M. Dussiot, T. T. Maciel, A. Fricot, *et al.*, "An activin receptor iia ligand trap corrects ineffective erythropoiesis in -thalassemia," *Nature Medicine*, vol. 20, pp. 398–407, 4 Apr. 2014.

[438] M. D. Cappellini, V. Viprakasit, A. T. Taher, *et al.*, "A phase 3 trial of luspatercept in patients with transfusion-dependent -thalassemia," *New England Journal of Medicine*, vol. 382, pp. 1219–1231, 13 Mar. 2020.

[439] K. Hestdal, S. E. Jacobsen, F. W. Ruscetti, D. L. Longo, T. C. Boone, and J. R. Keller, "Increased granulopoiesis after sequential administration of transforming growth factor-beta 1 and granulocyte-macrophage colony-stimulating factor.," *Experimental hematology*, vol. 21, pp. 799–805, 6 Jun. 1993.

[440] A. Celada and R. A. Maki, "Transforming growth factor-beta enhances the m-csf and gm-csf-stimulated proliferation of macrophages.," *Journal of immunology (Baltimore, Md. : 1950)*, vol. 148, pp. 1102–5, 4 Feb. 1992.

[441] T. A. Borkowski, J. J. Letterio, A. G. Farr, and M. C. Udey, "A role for endogenous transforming growth factor 1 in langerhans cell biology: the skin of transforming growth factor 1 null mice is devoid of epidermal langerhans cells," *The Journal of Experimental Medicine*, vol. 184, pp. 2417–2422, 6 Dec. 1996.

[442] O. Takeuchi and S. Akira, "Pattern recognition receptors and inflammation," *Cell*, vol. 140, pp. 805–820, 6 Mar. 2010.

[443] F. Martín-Sánchez, C. Diamond, M. Zeitler, *et al.*, "Inflammasome-dependent il-1 release depends upon membrane permeabilisation," *Cell Death Differentiation*, vol. 23, pp. 1219–1231, 7 Jul. 2016.

[444] A. Baroja-Mazo, F. Martín-Sánchez, A. I. Gomez, *et al.*, "The nlrp3 inflammasome is released as a particulate danger signal that amplifies the inflammatory response," *Nature Immunology*, vol. 15, pp. 738–748, 8 Aug. 2014.

[445] Z. Cao, J. Xiong, M. Takeuchi, T. Kurama, and D. V. Goeddel, "Traf6 is a signal transducer for interleukin-1," *Nature*, vol. 383, pp. 443–446, 6599 Oct. 1996.

[446] T. Liu, L. Zhang, D. Joo, and S.-C. Sun, "Nf-b signaling in inflammation," *Signal Transduction and Targeted Therapy*, vol. 2, p. 17 023, 1 Jul. 2017.

[447] E. M. Pietras, C. Mirantes-Barbeito, S. Fong, *et al.*, "Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal.," *Nature cell biology*, vol. 18, pp. 607–18, 6 Jun. 2016.

[448] J. S. Chavez, J. L. Rabe, D. Loeffler, *et al.*, "Pu.1 enforces quiescence and limits hematopoietic stem cell expansion during inflammatory stress," *Journal of Experimental Medicine*, vol. 218, 6 Jun. 2021.

[449] J. McClatchy, R. Strogantsev, E. Wolfe, *et al.*, "Clonal hematopoiesis related tet2 loss-of-function impedes il1-mediated epigenetic reprogramming in hematopoietic stem and progenitor cells.," *Nature communications*, vol. 14, p. 8102, 1 Dec. 2023.

[450] K. Liang, A. G. Volk, J. S. Haug, *et al.*, "Therapeutic targeting of mll degradation pathways in mll-rearranged leukemia," *Cell*, vol. 168, 59–72.e13, 1-2 Jan. 2017.

[451] T. W. Combes, F. Orsenigo, A. Stewart, *et al.*, "Csf1r defines the mononuclear phagocyte system lineage in human blood in health and covid-19," *Immunotherapy Advances*, vol. 1, 1 Jan. 2021.

[452] D. A. Baker, H. R. Glover, and N. J. Dibb, "Synergy between scf or m-csf with il-3 or gm-csf in fdc-p1 cells: A sensitive assay of transforming mutations of c-fms.," *Leukemia*, vol. 8, pp. 141–50, 1 Jan. 1994.

[453] F. Bot, L. van Eijk, L. Broeders, L. Aarden, and B. Lowenberg, "Interleukin-6 synergizes with m-csf in the formation of macrophage colonies from purified human marrow progenitor cells," *Blood*, vol. 73, pp. 435–437, 2 Feb. 1989.

[454] E. Kim, J. O. Ilagan, Y. Liang, *et al.*, "Srsf2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition," *Cancer Cell*, vol. 27, pp. 617–630, 5 May 2015.

[455] A. A. Basiorka, K. L. McGraw, E. A. Eksioglu, *et al.*, "The nlrp3 inflammasome functions as a driver of the myelodysplastic syndrome phenotype.," *Blood*, vol. 128, pp. 2960–2975, 25 Dec. 2016.

[456] M. Gozdecka, M. Dudek, S. Wen, *et al.*, "Mitochondrial metabolism sustains dnmt3a-r882-mutant clonal haematopoiesis," *Nature*, vol. 642, pp. 431–441, 8067 Jun. 2025.

[457] T. Pabst, B. U. Mueller, P. Zhang, *et al.*, "Dominant-negative mutations of cebpa, encoding ccaat/enhancer binding protein- (c/ebp), in acute myeloid leukemia," *Nature Genetics*, vol. 27, pp. 263–270, 3 Mar. 2001.

[458] B. A. Caldwell, Y. Wu, J. Wang, and L. Li, "Altered dna methylation underlies monocyte dysregulation and immune exhaustion memory in sepsis," *Cell Reports*, vol. 43, p. 113 894, 3 Mar. 2024.

[459] H. Pimentel, M. Parra, S. Gee, *et al.*, "A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis," *Nucleic Acids Research*, vol. 42, pp. 4031–4042, 6 Apr. 2014.

[460] J. J. Xu, A. M. Chalk, M. Wall, W. Y. Langdon, M. F. Smeets, and C. R. Walkley, "Srsf2p95h/+ co-operates with loss of tet2 to promote myeloid bias and initiate a chronic myelomonocytic leukemia-like disease in mice," *Leukemia*, vol. 36, pp. 2883–2893, 12 Dec. 2022.

[461] L. Hurtado-Navarro, E. J. Cuenca-Zamora, L. Zamora, *et al.*, "Nlrp3 inflammasome activation and symptom burden in kras-mutated cmml patients is reverted by il-1 blocking therapy.," *Cell reports. Medicine*, vol. 4, p. 101 329, 12 Dec. 2023.

[462] M. Askmyr, H. Ågerstam, N. Hansen, *et al.*, "Selective killing of candidate aml stem cells by antibody targeting of il1rap," *Blood*, vol. 121, pp. 3709–3713, 18 May 2013.

[463] M. Järås, P. Johnels, N. Hansen, *et al.*, "Isolation and killing of candidate chronic myeloid leukemia stem cells by antibody targeting of il-1 receptor accessory protein," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 16 280–16 285, 37 Sep. 2010.

[464] K. A. West, S. Kumar, M. Scott, *et al.*, "Bos-371, a monoclonal antibody against il1rap: Characterization in preclinical models of aml.," *Journal of Clinical Oncology*, vol. 41, pp. 7030–7030, $16_{s}uppl$ Jun. 2023.

[465] Y. Zhang, M. Park, L. Y. Ghoda, *et al.*, "Il1rap-specific t cell engager depletes acute myeloid leukemia stem cells," *Journal of Hematology Oncology*, vol. 17, p. 67, 1 Aug. 2024.

[466] D. J. Glass and N. Hall, "A brief history of the hypothesis," *Cell*, vol. 134, pp. 378–381, 3 Aug. 2008.

[467] T. Ifargan, L. Hafner, M. Kern, O. Alcalay, and R. Kishony, "Autonomous llm-driven research — from data to human-verifiable research papers," *NEJM AI*, vol. 2, 1 Jan. 2025.

[468] X. Gu, D. Jiang, S. Schlanger, *et al.*, "Metabolic signatures and pathway rewiring in tet2-mutated myeloid neoplasia," *Blood*, vol. 142, pp. 2747–2747, Supplement 1 Nov. 2023.

[469] A. O. Khan, A. Rodriguez-Romera, J. S. Reyat, *et al.*, "Human bone marrow organoids for disease modeling, discovery, and validation of therapeutic targets in hematologic malignancies," *Cancer Discovery*, vol. 13, pp. 364–385, 2 Feb. 2023. [Online]. Available: `https://doi.org/10.1158/2159-8290.CD-22-0199`.

[470] S. G. Cockey, H. Zhang, M. O. Hussaini, *et al.*, "A large cohort study of 412 patients with *SRSF2/TET2* co-mutated myeloid neoplasms: The molecular landscape and clinical outcomes," *Blood*, vol. 142, pp. 1882–1882, Supplement 1 Nov. 2023.

[471] C. Haferlach, A. Stengel, M. Meggendorfer, W. Kern, and T. Haferlach, "Characterization of mds harboring tet2 mutations and/or tet2 deletions," *Blood*, vol. 128, pp. 4288–4288, 22 Dec. 2016.

[472] Z. Zou, X. Dou, Y. Li, *et al.*, "Rna m5c oxidation by tet2 regulates chromatin state and leukaemogenesis," *Nature*, vol. 634, pp. 986–994, 8035 Oct. 2024.