Modelling GATA2 Immunodeficiency and the progression towards Myelodysplastic syndrome (MDS) and Acute myeloid leukaemia (AML)

Leigh-anne Thomas

A thesis submitted for the Doctor of Philosophy at

CARDIFF UNIVERSITY

School of Biosciences

Submitted: March 2021

Leigh-anne Thomas is supported by Leukaemia Research Appeal Wales

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Neil Rodrigues for his continued support and guidance throughout these years. I would like to express my immense gratitude for giving me the opportunity to embark on this PhD.

I would also like to thank Dr. Fernando Afonso Dos-Anjos, Professor Alex Tonks and Dr Gui Jie Feng for their training, guidance and helpful advice. I would like to thank past and present colleagues of the Rodrigues' lab, in particular Hind Alqahtani, Dr Ali Abdelfattah and Sarab Taha for their guidance and assistance during the mouse experiments. A big thank-you to Jolene Twomey for your kind words always offered, without you the institute would not function.

A very special thanks goes to Dr Juan Bautista Menéndez González for making me strive to be a better researcher. You have inspired me throughout the PhD with your work ethic and passion for science. I have no doubt you will be a very successful group leader in the future. Thank-you for your help and encouragement during the testing times.

I would like to thank my colleagues and friends at the European Cancer Stem Cell Research Institute for creating a working environment that has never quiet felt like work. In particular, Dr Ana Jimenez Pascual who was always up for chat and procrastination and Dr Carlotta Olivero for her infectious positive attitude. A special thanks to Dr Giusy Tornillo for her vast knowledge and wise advice throughout the challenges of this PhD; you are a very good friend, and I will miss working alongside you.

A very special thanks to my family, for their endless love and support during these years, some of which have been a challenge, thank-you for helping me through. A special mention to my grandparents David Thomas, Sheila McGrath and Michael McGrath who supported me financially during my Masters, without which I would not have had the opportunity to continue with this PhD. You all have played an instrumental role in progressing my career, I am truly grateful.

Last but not least, I would like to thank Leukaemia Appeal Wales for my studentship and for supporting my research. Without you it would not have been possible.

Summary

Heterozygous germline mutations in the GATA2 gene gives rise to an immunodeficiency syndrome characterised by cytopenias, severe infections and a predisposition towards haematological maligancies such as myeloidysplastic syndrome (MDS) and acute myeloid leukemia (AML). Clinical data from GATA2 mutated families reveal a variable age at presentation, a variable disease phenotype and a proportion of asymptomatic carriers, suggesting incomplete penetrance. Acquisition of secondary hits such as somatic mutations in the ASXL1 gene and cytogenetic abnormalities such as monosomy 7 and trisomy 8 are commonly found in patients that have progressed towards MDS/AML. Perturbation in the immune system of these patients leads to persistent inflammation. Indeed, scRNA-seq from patient samples showed immune and inflammatory pathway dysregulation at the HSC level, which has also been confirmed in our lab in HSCs from *Gata2* heterozygous mice. Whilst there has been much research over the years investigating the loss of GATA2 in a variety of mouse and human models, many have failed to recapitulate key aspects of the disease. Using human iPSCs, primary cord blood CD34+ cells and transgenic mouse models, GATA2 immunodeficiency was modelled along with further investigations into inflammation as a potential driver of disease progression.

In this thesis, CRISPR/Cas9 and CRISPR base editors were used to engineer specific GATA2 mutations - T354M and R361C – into human iPSC lines followed by truncating ASXL1 mutations to model GATA2 immunodeficiency and its progression towards myeloid malignancy. Unfortunately, due to technical and time limitations (including COVID19) this work was unable to be taken to fruition. Interestingly, *in vitro* bacterial stimulation selectively increased CFU output and CD14 expression in GATA2 knockdown cord blood $CD34^{+}$ cells. In addition, chronic inflammation led to splenomegaly and granuloma formation along with a reduction of $Gr1^{+}Mac1^{+}$ cells in the BM of mice with haematopoietic-specific deletion of a single *Gata2* allele. Finally, in comparison with steady-state conditions where *Gata2*-heterozygous HSC levels were reduced by half, HSC numbers remained unchanged while control HSCs were reduced by half upon inflammation, indicating that *Gata2* heterozygosity enabled HSCs mechanism to better adapt to inflammation. Taken together, this work highlights that a multi-model approach is desirable to study GATA2 familial MDS/AML and that inflammation is an important mediator for disease progression in GATA2 immunodeficiency syndromes thus opening new avenues for disease modelling and therapeutic intervention.

Table of contents

Acknowledgements ..... **II**

Summary ..... **III**

Table of contents ..... **IV**

List of tables ..... **VII**

List of figures ..... **VIII**

Abbreviations ..... **XI**

Chapter 1 ..... **1**

Introduction **1**

1.1 Haematopoiesis 2
1.1.1 Haematopoietic development 2
1.1.2 Adult Haematopoiesis 3
1.1.3 Bone marrow niche 7
1.1.4 Regulatory transcription factors of haematopoiesis 13

1.2 Myeloid malignancies 26
1.2.1 Myelodysplastic syndrome (MDS) 26
1.2.2 Acute myeloid leukaemia (AML) 31
1.2.3 Classification of myeloid malignancies 37
1.3 Myeloid neoplasms with a germline predisposition 40
1.3.4 GATA2 myeloid malignancy 45

1.4 Aims and objectives of thesis 69

Chapter 2 ..... **70**

Materials and Methods **70**

2.1 Cell lines 71

2.2 Human iPS cell culture 73
2.2.1 Essential 8 culture system 73
2.2.2 StemFlex culture system 73

2.3 Cryopreservation of cell lines 73

2.4 RNA extraction and gene expression analysis 74

2.5 Immunofluorescence 74

2.6 Chromosome quality assessment 75

2.7 Human iPSC differentiation 75
2.7.1 Human iPSC differentiation in OP9 co-culture 75
2.7.1.1AutoMACS magnetic cell separation- CD34 Enrichment 76
2.7.2 STEMdiff™ Haematopoietic Differentiation 76
2.7.3 Colony Forming Unit (CFU) Assay 76

2.8 CRISPR/CAS9 editing 77
2.8.1 Forming CRISPR/CAS9 RNP complex 77
2.8.2 Electroporation 77
2.8.3 DNA extraction and Genomic PCR 77
2.8.4 Surveyor Assay 78
2.9 CRISPR/Cas9 Base editing 78
2.9.1 DNA vectors and preparation 79
2.9.2 U6gRNA Cloning 79
2.9.3 Base editing transfection 81
2.9.4 TaqMan SNP Genotyping Assay 81
2.10 shRNA mediated KD in CB $CD34^+$ cells 83
2.10.1 Viral Plasmids 83
2.10.2 Generation of lentiviral particles 83
2.10.3 Titration of lentiviral particles 86
2.10.4 Primary Cord Blood cell isolation 86
2.10.5 CD34 manual enrichment 87
2.10.6 Lentiviral transduction of $CD34^+$ HPSC 88
2.10.7 Protein isolation and cell lysate preparation 88
2.10.8 Protein quantification 88
2.10.9 Protein separation and transfer to blot 89
2.10.10 Protein blotting and visualisation 89
2.10.11 Human CFU Assay 90
2.10.12 $CD34^+$ CB myeloid expansion with LPS challenge 90
2.10.13 Detection of apoptosis by annexin V 90
2.11 Mice 91
2.11.1 Gata2 mouse strain ($Gata2^{tm1Sac}$) 91
2.11.2 Vav-iCre breeding with $Gata2^{tm1Sac}$ and induction of Gata2 deletion in vivo. 92
2.11.3 DNA extraction and genotyping of animals 92
2.11.4 Induction of Polyinosinic:polycytidylic acid (poly I:C) 93
2.11.5 Dissection and PB extraction 93
2.11.6 Red blood lysis 93
2.11.7 Bone marrow extraction 94
2.11.8 Splenocyte extraction 94
2.11.9 Murine Colony forming cell (CFC) assay 94
2.12 Flow Cytometry 94
2.12.1 Intracellular staining 94
2.12.2 Extracellular staining 95
2.13 Cell sorting 95
2.13.1 Haematopoietic Progenitor Cell (HPC) sorting 95
2.13.2 Single cell sorting into 96 wells 96
2.14 Flow cytometry Analysis 96
2.15 Data and statistical Analysis 96
Chapter 3 99
Characterisation of human iPS cell lines 99
3.1 Introduction 100

3.1.1 iPSC technology 100
3.1.2 PSC haematopoietic differentiation 102
3.2 Aims and objectives 107
3.3 Results 108
3.3.1 Characterisation of parental human iPS cell lines. 108
3.3.2 Optimising human iPS haematopoietic differentiation 111
3.4 Discussion 123
Chapter 4: 126
Modelling patient specific GATA2 mutations using CRISPR/Cas9 technology 126
4.1 Introduction 127
4.1.1 Human iPSC disease modelling 127
4.1.2 CRISPR-Cas genome editing tools 128
4.2 Aims and objectives 136
4.3 Results 137
4.3.1 Generating heterozygous GATA2 T354M mutation using CRISPR-Cas9 nucleases 137
4.3.2 Generate heterozygous GATA2 T354M and GATA2 R361C mutations using CRISPR-Cas9 Base editors. 144
4.4 Discussion 153
Chapter 5: 157
Exploring the role of inflammation as a driver of disease progression during early stages of Gata2 syndromes 157
5.1 Introduction 158
5.2 Aims & Objectives 162
5.3 Results 163
5.3.1 Validation of shRNA vectors against human GATA2 in CB HPSCs. 163
5.3.2 LPS stimulation in GATA2 KD cells results in an increased progenitor output. 164
5.3.3 GATA2 KD cells have reduced late phase apoptosis 165
5.3.4 CD14 is highly expressed in GATA2 KD cells stimulated with LPS. 168
5.3.5 pIpC stimulation in Gata2 heterozygous mice results in decreased myeloid output in the BM. 168
5.3.5 pIpC stimulation in Gata2 heterozygous mice results in splenomegaly. . 170
5.3.6 pIpC exposure in Gata2 heterozygous mice has no effect on stem and progenitor output in the BM. 177
5.3.7 pIpC exposure in Gata2 heterozygous mice has no effect mature cells in the spleen. 181
5.3.8 Chronic pIpC stimulation in Gata2 haploinsufficient mice does not impact the DNA damage response. 187
5.4 Discussion 189
Chapter 6: 195
General Discussion 195

6.1 Towards a more specific model of GATA2 immunodeficiency 196
6.2 Inflammation is an important factor in GATA2 syndrome disease progression 199
6.3 Molecular and biological implications for GATA2 immunodeficiency- What do we know? 201
6.4 Towards better clinical management of GATA2 syndromes 202

List of tables

Table 1.1: GATA2 Cis-acting sites and function ……..……..……21
Table 1.2: MDS chromosomal abnormalities and their frequencies at diagnosis ……..……27
Table 1.3: Therapies for the treatment of MDS ……..……30
Table 1.4: Molecular events in AML leukemogenesis ……33
Table 1.5: Prognosis- risk categories of AML based on cytogenetic profile ……..34
Table 1.6: Therapeutic agents for treatment of AML ….……..……..36
Table 1.7a: French-American-British (FAB) classification of MDS ……..37
Table 1.7b: French-American-British (FAB) classification of AML ……..……37
Table 1.8: MDS Myeloid neoplasms WHO classifications ….……38
Table 1.9: Acute myeloid leukaemia (AML) and related neoplasms ……39
Table 1.10: Predisposing haematopoietic transcription factors in familial myeloid neoplasm……..……..40
Table 1.11: GATA2 immunodeficiency syndrome ……..……48
Table 1.12: Clinical heterogeneity reported among GATA2 deficient patients ….……..……49
Table 1.13: Prevalence of GATA2 deficiency related myeloid neoplasms …….……52
Table 1.14: GATA2 patient monitoring and screening ……54
Table 2.1: RT-qPCR primers …….……74
Table 2.2: Sequences for CRISPR/CAS9 editing of the GATA2 region T354 ……77
Table 2.3: Genomic PCR primers and corresponding annealing conditions ……..78
Table 2.4: U6 gRNA oligo sequenc……79
Table 2.5: shRNA sequences GATA2, exon location following ……..…….83
Table 2.6: Pierce™ BCA Protein Assay standard and sample preparation ……...…89
Table 2.7: Primary and secondary antibodies used during western blotting protocol ….…..90
Table 2.8: Genotyping primers and cycling conditions ..…….……....93
Table 2.9: Mouse antibodies used for flow cytometry (FC) …….……..……97
Table 2.10: Human antibodies used for flow-cytometry (FC) and immunofluorescence (IF) ……..98
Table 3.1: Delivery methods of reprogramming factors for induced pluripotent stem (iPS) cell generation
……..……..102
Table 4.1: Base editors used to correct or model mutations ……...134

List of figures

Figure 1.1: Murine Classical Haematopoietic hierarchy……..……4
Figure 1.2: Human haematopoietic hierarchy ..……7
Figure 1.3: The bone marrow niche ……..……9
Figure 1.4: GATA2 isoform and protein regulation ..……20
Figure 1.5: Germline and sporadic GATA2 mutations ……45
Figure 1.6: Lineages documented to be affected in patients harbouring GATA2 germline mutations ……53
Figure 1.7: ASXL1 wild-type and mutant protein ……59

Figure 2.1: U6gRNA- cloning approach ……80
Figure 2.2: Vector map of plasmids used in gene editing experiments ……..…….82
Figure 2.3: Vector map of psi-LVU6GP plasmids used in shRNA experiments. ……85
Figure 2.5: Annexin V detection of apoptosis ……..……..…….91
Figure 2.6: Strategy used for the engineering of the Gata2^{tm1sac} transgenic mouse model …….. 92
Figure 2.7: Gating strategy used for Haematopoietic differentiation of pluripotent stem (PSC) cells using the StemDiff culture system ……..……..…….95
Figure 2.8: Gating strategy for single cell sorting of TRA-1-60 positive human iPS cells …96
Figure 3.1: Stages of haematopoietic development from human PSCs ……..……..…..104
Figure 3.2: Human iPS cell lines EPI and KOLF2 carry the hallmarks of pluripotency …. 109
Figure 3.3: Human cell lines used as positive and negative controls for pluripotency analysis ……....110
Figure 3.4: Human PSC lines present with a euploid karyotype according to metaphase counts …..……..……111
Figure 3.5: Human pluripotent stem cell haematopoietic differentiation using the OP9 co-culture system ……112
Figure 3.6: Phenotypic analysis of EPI-iPSC initiated on OP9 co-cultures with varying conditioning phases ……113
Figure 3.7: Phenotypic analysis of EPI-iPSC seeded at a higher density to initiate OP9 co-cultures varying OP9 conditioning phases ..……..…114
Figure 3.8: Phenotypic analysis of KOLF2-iPSC haematopoietic differentiation using the OP9 co-culture system with varying conditioning phases ……..…..……115
Figure 3.9: Human iPSC haematopoietic differentiation using the STEMdiff™ culture …..116
Figure 3.10: The human EPI-iPS cell line efficiently differentiates into HPCs using the STEMdiff™ culture system ……..……118

Figure 3.11: Kinetic analysis of haematopoietic development in human iPS-EPI line ..…119
Figure 3.12: The EPI-iPSC line efficiently produces HPCs and has high CFU potential with multilineage capacity ……120
Figure 3.13: Kinetic analysis of haematopoietic development in human iPS-KOLF2 line .121

Figure 4.1: Cas9 Complex binding of target DNA ……129
Figure 4.2: Gene editing using CRISPR/Cas9 ……130
Figure 4.3: Third- generation base editor (BE3) facilitates specific guide RNA programmed C.G >T.A modification ……133
Figure 4.4: Generating human iPSC lines with GATA2 T354M heterozygous mutation using CRISPR/Cas9 genome editing . ……138
Figure 4.5: Schematic illustration of experimental plan ……139
Figure 4.6: Screening CRISPR/Cas9 clones using the Surveyor cleavage assay ……141
Figure 4.7: Mutations incorporated at the T354 site in human iPS cells using CRISPR/Cas9 editing ……..143
Figure 4.8: Base editing design of GATA2 mutations T354M and R361C ……146
Figure 4.9: Base editing design of secondary mutations in ASXL1; G1300X and R1415X ……148
Figure 4.10: Base editing experiment workflow ……149
Figure 4.11: Screening of clones using the TaqMan SNP genotyping assay ……150
Figure 4.12: Delivery of BE4-MAX GFP plasmid by electroporation ……151

Figure 5.1: Knockdown validation of human short hairpin constructs against GATA2 ..…163
Figure 5.2: Colony forming cell assay of GATA2-KD or control $CD34^+$ CB cells stimulated with LPS ……164
Figure 5.3: LPS stimulation in GATA2 KD cells results in an increased progenitor output ……..165
Figure 5.4: Myeloid differentiation of GATA2 KD or control $CD34^+$ CB cells stimulated with LPS ……...166
Figure 5.5: LPS stimulation in GATA2 KD cells results in reduced cell number during myeloid differentiation and a GATA2 mediated effect on late apoptosis ……...167
Figure 5.6: CD14 is highly expressed in *GATA2* KD cells stimulated with LPS ……168
Figure 5.7: pIpC stimulation in Gata2 heterozygous mice results in splenomegaly ……..170
Figure 5.8: *Gata2* heterozygosity during chronic pIpC stimulation in mice has no effect on mature lymphoid cells in the BM ……..172

Figure 5.9: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on multipotent monocyte subsets in the BM ……..174
Figure 5.10: Gating strategy for immature myeloerythroid cell lineages in the bone marrow during inflammation ……..175
Figure 5.11: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on immature myeloerythroid cell lineages in the bone marrow ……..…176
Figure 5.12: CFC and replating assays from total BM cells from pIpC stimulated *Gata2* heterozygous mice ……..177
Figure 5.13: Gating strategy for multipotent progenitor and common lymphoid progenitor compartments in the bone marrow during inflammation ……..…..179
Figure 5.14: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on multipotent progenitor and common lymphoid progenitor populations in the BM ……180
Figure 5.15: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on mature populations in the spleen ……..……..182
Figure 5.16: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on multipotent monocyte subsets in the Spleen ……183
Figure 5.17: Chronic pIpC exposure in *Gata2* heterozygous mice has no impact on progenitor populations in the spleen ……184
Figure 5:18: Gating strategy for stem and progenitor cells in the spleen ……185
Figure 5.19: Chronic pIpC exposure in *Gata2* heterozygous mice has no effect on multipotent progenitor and common lymphoid progenitor populations in the spleen ……...186
Figure 5.20: *Gata2* heterozygosity during chronic pIpC exposure in mice leads to an increase in myeloid precursors in the PB ……..……..187
Figure 5.21: *Gata2* heterozygosity during chronic pIpC exposure in mice shows no effect of phosphorylation of γ-H2AX in the BM ……..188
Figure 5.22: Modelling Gata2 haploinsufficiency using the *Gata2* Vav-iCe mouse model 193
Figure 5.23: Modelling *Gata2* haploinsufficiency using the *Gata2* Vav-iCe mouse model ……..……..194

Abbreviations

‘ = Minutes

2-HG = 2-Hydroxygluterate

5-FU = 5’-fluorouracil

6-OHDA = 6-Hydroxydopamine

AGM = Aorta-gonad mesonephros

ALL = Acute lymphoblastic leukaemia

AML = Acute Myeloid Leukaemia

AML1/ETO= Acute Myeloid Leukaemia/ Eight-Twenty One oncoprotein

Ang-1 = Angiopoietin-1

APL = Acute Promyelocytic Leukaemia

Ara-c = Cytarabine

BFU-E = Burst Forming Unit-Erythrocyte

BM = Bone marrow

BMP = Bone Morphogenic Protein

BMT = Bone marrow transplant

bp = base pair

CAR = Cxcl12-abundant reticular cells

Cas= CRISPR associated protein

CB = Cord blood

cDNA = Complementary DNA

CEBPA = CCAAT/enhancer binding protein alpha

CFC = Colony Forming Cell

CFU = Colony Forming Unit

CFU-E = Colony Forming Unit- Erythrocyte

CFU-G = Colony Forming Unit- Granulocyte

CFU-GEMM = Colony Forming Unit- Granulocyte/Erythrocyte/Macrophage/Megakaryocyte

CFU-GM = Colony Forming Unit- Granulocyte/Macrophage

CFU-M = Colony Forming Unit- Macrophage

CFU-Meg/Mk = Colony Forming Unit- Megakaryocyte

CFU-S = Colony Forming Unit-Spleen

CGaP= Cellular Genotyping and Phenotyping facility

Cited2= CBP/p300 interacting transactivators with glutamic acid (E) and aspartic acid (D)-rich tail 2.

CLP= Common lymphoid progenitor CML = Chronic myeloid leukaemia

CMP= Common myeloid Progenitor

Cre= Cre recombinase

CRISPR= Clustered Regularly Interspaced Palindromic Repeats

CRISPRa= CRISPR activation

CRISPRi= CRISPR interference

crRNA= CRISPR RNA

CT= Cycle threshold

dCas9= deactivated/dead Cas9

DMSO= Dimethyl sulfoxide

DNA= Deoxyribonucleic Acid

DNase= Deoxyribonuclease

DNMT= DNA Methyltransferase

dNTP= Deoxynucleotide triphosphate

DSB= double-strand break

EDTA= Ethylenediaminetetraacetic acid

EGF= Epidermal Growth Factor

EGFR= Epidermal Growth Factor Receptor

EpoR= Erythropoietin receptor

Evi1= Ecotropic proviral integration site 1

FAB= French-American-British classification

FACS= Fluorescence Activated Cell Sorting

FBS= Foetal Bovine Serum

Flt3= FMS-like tyrosine kinase 3

g= Gram

GATA= GATA binding protein

G-CSF= Granulocyte-colony Stimulating Factor

Gfi1= Growth Factor independent 1

GFP= Green Fluorescent Protein

GMP= Granulocyte Macrophage Progenitor

GVHD= Graft versus Host Disease

gRNA= guide RNA

Hif= Hypoxia inducible factor

HLA = Human Leukocyte Antigen

Hox = Homeobox

HPC = Haematopoietic Progenitor Cell

HSC = Haematopoietic Stem Cell

HSPC = Haematopoietic Stem and Progenitor Cell

HDR= homology-directed repair

hESCs= human embryonic stem cells

IDH1/2 = Isocitrate dehydrogenase1/2

IL = Interleukin

Indel= insertion/deletion

iPSC= induced pluripotent stem cells

IRES = Internal-ribosome-entry-site

KO= knockout

kg = Kilogram

L = Litre

LB= Luria broth

Lin = Lineage

LMPP = Lymphoid-primed multipotent progenitors

LoxP = Locus of crossover of Bacteriophage P1

LSC = Leukaemia Stem Cell

LSK = Lineage-Sca-1^{+}c-Kit^{+}

LT-HSC = Long Term- HSC

LTC-IC = Long term culture initiating cell

MDS = Myelodysplastic syndrome

Meis = Myeloid ectropic insertion site

MEP = Megakaryocyte Erythrocyte Progenitor

mg = Milligram

mL = Millilitre

MLL = Mixed Lineage Leukaemia

mM = Millimolar

MPN = Myeloproliferative neoplasm

MPP = Multipotent Progenitor

mRNA = Messenger Ribonucleic Acid

MSC = Mesenchymal Stem Cell

MSCV = Murine stem cell virus

N = Number of biological replicates

NOD/SCID = Non-obese diabetic/severe combined immune deficiency

MEFs= mouse embryonic fibroblasts

MOI= multiplicity of infection

MUT=mutant

NHEJ= non-homologous end joining

OB = Osteoblasts

PAM= protospacer-adjacent motif

PBS = Phosphate Buffered Saline

PCR = Polymerase Chain Reaction

pIpC = polyinosinic:polycytidylic acid

qRT-PCR = Quantitative Reverse Transcription Polymerase Chain Reaction

RBC = Red blood cells

RNA = Ribonucleic Acid

RNase = Ribonuclease

ROS = Reactive oxygen species

rpm = Revolutions per minute

RT = Room Temperature

Runx1 = Runt-related transcription factor 1

SA = Streptavidin

Scf = Stem cell factor

Scl/Tal1 = Stem cell leukaemia gene

SEM = Standard error of mean

shRNA= short hairpin RNA

siRNA short interfering RNA

ST-HSC = Short term-HSCs

TBI = Total Body Irradiation

TET2 = Ten-eleven-translocation gene 2

TF = Transcription factor

TGF-β = Transforming Growth Factor-β

TPO = Thrombopoietin

tracrRNA= trans-activating crRNA

UV = Ultra Violet

VEGF = Vascular Endothelial Growth Factor

VEGFR = Vascular Endothelial Growth Factor Receptor

WBM = Whole Bone Marrow

WHO = World health organisation

Wnt = Wingless

WT = Wild Type

YFP = Yellow Fluorescent Protein

Chapter 1

Introduction

1.1 Haematopoiesis

Haematopoiesis is a life-long process that supplies the body with all blood cell types which is continuously being replenished in response to daily demands. During adulthood, haematopoiesis occurs mainly in the bone marrow (BM), though in some situations such as severe infection, pregnancy and dramatic blood loss, it also takes place in extramedullary sites such as the spleen and liver (Orkin and Zon, 2008; Doulatov, Notta, Laurenti and J. E. E. Dick, 2012; Clapes, Lefkopoulos and Trompouki, 2016a).

1.1.1 Haematopoietic development

Haematopoiesis is initiated in the mammalian embryo in three distinct programs. The first wave is initiated in the yolk sac (YS) blood islands at embryonic day 7 (E7). The first cohort of circulating 'primitive' red blood cells are derived from a transient population of extra-embryonic yolk sac progenitors, temporarily associated with megakaryocyte and macrophage potential, and are distinguished from their mature counterpart by their; large size, distinct expression of embryonic globin (βH1) and the presence of a nucleus (McGrath *et al.*, 2015; Kauts, Vink and Dzierzak, 2016). Primitive macrophages which do not go through a monocyte precursor are also derived from the blood islands during this stage of development (E7-7.5). Once the blood stream has been established at E8.25-85 these macrophages migrate to developing tissues where they become 'tissue resident' macrophages. Shortly before birth these are replaced by monocyte derived cells generated later in development (Kauts, Vink and Dzierzak, 2016).

The second wave also occurs in the YS at embryonic day 8.25 (E8.5) and overlaps with the primitive stage. Functionally more complex bipotential erythro-myeloid progenitors (EMPs) which colonize the fetal liver (FL) at E9 giving rise to large numbers of erythrocytes, macrophages and granulocytes. Erythrocytes derived from these EMPs are distinct from those found in the primitive wave as they express adult (βmajor) goblin and undergo enucleation. Moreover, EMPs are only able to provide short term *in vivo* reconstitution and lack lymphoid potential (Frame, McGrath and Palis, 2013; McGrath *et al.*, 2015). Other haematopoietic cells generated in the YS during this wave are B-1 B cell progenitors (E8.5-9.5) and mast cells at E9.5. It is believed that the generation of these more adult cells in this phase of development plays an important role in promoting hemogenesis during the final wave (Kauts, Vink and Dzierzak, 2016).

The final intraembryonic definitive stage of haematopoiesis occurs in the embryonic arterial vascular in the aorta-gonas-mesonephoros (AGM). Herein, under a process termed the 'endothelial to haematopoietic transition' (EHT), specialised endothelial cells- termed hemogenic endothelial cells (HEC), give rise to the first HSC (E10.5). Additionally, HSCs are

found in significant numbers in the placenta (E11) and increase for several days after (Gekas *et al.*, 2005; Ottersbach and Dzierzak, 2005). Whilst the YS and placenta are sites of HSC generation, the FL serves as a niche for the expansion of HSCs and EMPs generated by other tissues (Kauts, Vink and Dzierzak, 2016). Just before birth, HSC migrate to the BM where they reside throughout adult life in specialised niches.

1.1.2 Adult Haematopoiesis

The adult haematopoietic system consists of a hierarchy of cells progressing from a rare population of long lived, self-renewing haematopoietic stem cell (HSC) that give rise to more terminally differentiated progenitor cells over various blood lineages. As mature blood cells are short lived, HSCs play a critical role throughout an individuals' life by replenishing multilineage progenitors and committed precursors to yield mature blood cells; red blood cells, megakaryocytes, myeloid cells and lymphocytes (Orkin and Zon, 2008). As you move from the apex of the hierarchy, these progenitors lose their multilineage and self-renewing potency before becoming fully mature functional blood cells (Kauts, Vink and Dzierzak, 2016).

1.1.2.1 Murine hierarchy

The identification of multiple surface markers such as CD34, Sca1 and the SLAM family receptors over the last few decades, along with the advancement of Fluorescence-activated cell sorting (FACS) has enabled the isolation of defined stem and progenitor cell populations (Chen *et al.*, 2008; Oguro, Ding and Morrison, 2013). Moreover, transplantation assays in mice have been key tools in identifying the short term and long-term engraftment potential of HSC and progenitor populations, aiding our current understanding of the haematopoietic hierarchy (Herzenberg *et al.*, 2006; Purton and Scadden, 2007). The classical hierarchy describes a stepwise differentiation process from HSCs to their progenies. HSCs were reported to be classified into two subpopulations, the first long-term HSC (LT-HSCs) ($Lin^-Sca1^+c\text{-}kit^+FLK2^-CD150^+CD48^-$) are a rare population of BM cells which are quiescent under steady-state conditions. Once exposed to stress stimulus these cells reactivate, enter the cell cycle and are capable of reconstituting the haematopoietic system as shown in lethally irradiated animals (Morrison and Weissman, 1994; Schoedel *et al.*, 2016). The second population- short-term HSCs (SH-HSCs) ($Lin^-Sca1^+c\text{-}kit^+FLK2^-CD150^-CD48^-$) have short term reconstitution ability rapidly restoring the haematopoietic system for approximately 8-12 weeks post transplantation. These ST-HSCs differentiate into multipotent progenitors (MPPs) ($Lin^-Sca1^+c\text{-}kit^+CD150^-CD48^-$) which are unable to self-renew, though enter the cell cycle more frequently and have more robust differentiation activity (Morrison *et al.*, 1997). From the MPP population, cells bifurcate into common lymphoid progenitors (CLPs) ($Lin^-Sca1^{low}c\text{-}kit^{low}CD127^+$) that possess only lymphoid-restricted differentiation capacity or the common

myeloid progenitors (CMPs) (Lin$^-$Sca1$^-$c-kit$^+$CD34$^+$CD16/32$^-$) that are able to differentiation towards megakaryocyte/erythroid progenitors (MEPs) (Lin$^-$Sca1$^-$c-kit$^+$CD34$^-$CD16/32$^-$) and granulocyte/macrophage progenitors (GMPs) (Lin$^-$Sca1$^-$c-kit$^+$ CD34$^+$CD16/32$^+$) (Morrison *et al.*, 1997; Na Nakorn *et al.*, 2002) **(Figure 1.1).**

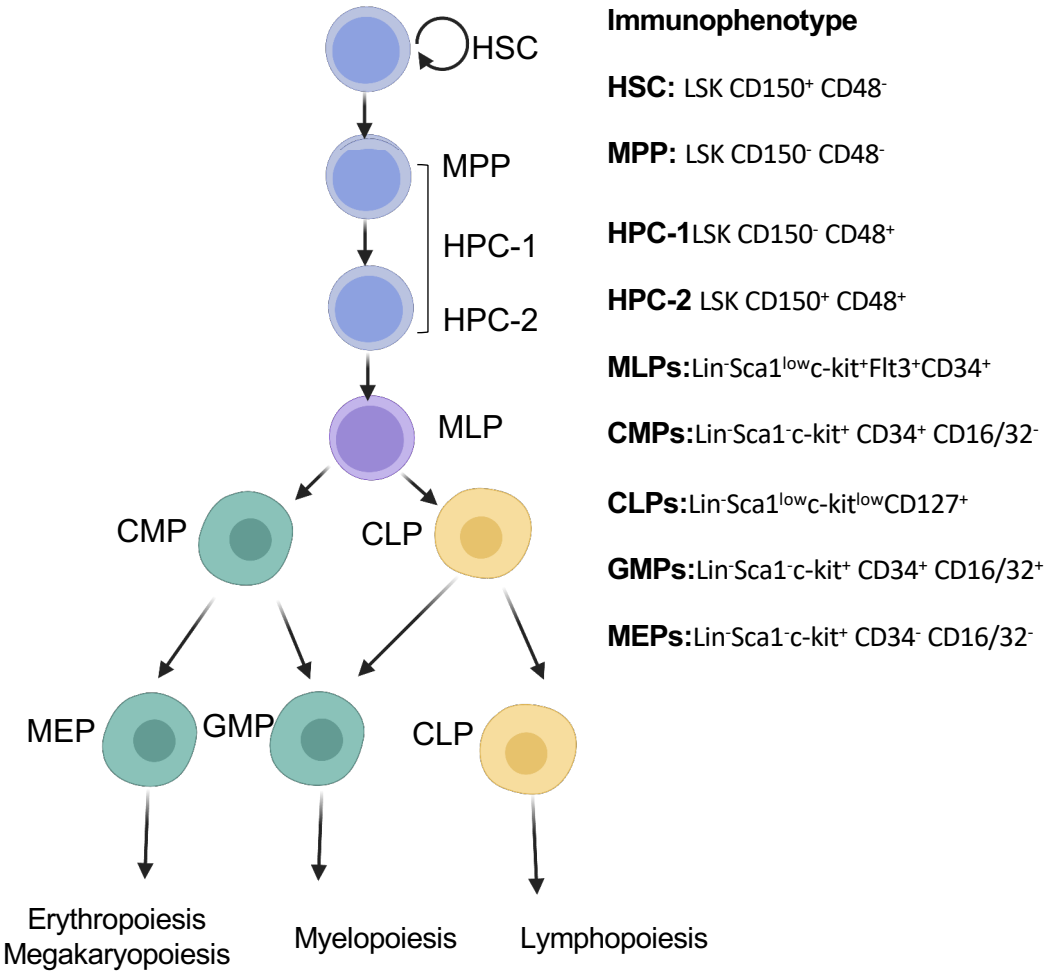


Figure 1.1. Murine Classical Haematopoietic hierarchy. HSCs reside at the top of the haematopoietic hierarchy where they sustain lifetime haematopoiesis. HSCs give rise to multipotent progenitor cells that progressively proliferative potential while differentiating into the main blood cell types.

Since this proposed classical system, studies have identified new haematopoietic cell populations revealing that the differentiation process is much more complex than previously thought, with several new revisions suggested. A new population of HSCs- intermediate-term HSCs (IT-HSCs) (Lin$^-$Sca1$^+$c-kit$^+$CD34^{lo}Flt3$^-$Rho^{lo}CD49b^{hi}) were identified based on their self-renewing ability, lying between LT (Lin$^-$Sca1$^+$c-kit$^+$CD34^{lo}Flt3$^-$Rho^{lo}CD49b^{lo}) and SH-HSCs (Lin$^-$Sca1$^+$c-kit$^+$CD34^{hi}Flt3$^+$Rho^{hi}) acting as a transitory population (Benveniste *et al.*, 2010; Yamamoto *et al.*, 2013). Furthermore, the MPP population was further subdivided into four

distinct populations; MPP1, MPP2 ($Lin^-Sca1^+c\text{-}kit^+$ $Flk2^-CD150^+CD48^+$), MPP3 ($Lin^-Sca1^+c\text{-}kit^+$ $Flk2^-CD150^-CD48^+$) and MPP4 ($Lin^-Sca1^+c\text{-}kit^+$ $Flk2^+CD150^-CD48^+$) based on their immunophenotype, BM abundance, cell-cycle status and their differentiation potential (Wilson *et al.*, 2008; Pietras *et al.*, 2015). The MPP1 population was proposed to be similar to the SH-HSCs, where MPP2-4 where defined as parallel subpopulations downstream of HSCs and are responsible for sustaining blood production during steady-state. Additionally, a subpopulation of HSCs with high expression of Flt3 was defined as lymphoid-primed multipotent progenitors (LMPPs), which preferentially generate lymphoid lineages, though possess the same immunophenotype as MPP4 (Adolfsson *et al.*, 2005; Cabezas-Wallscheid *et al.*, 2014; Pietras *et al.*, 2015).

Functional assays, lineage tracing and single cell transcriptome analysis has further challenging the way in which we visualise the haematopoietic hierarchy. According to these 'classical' models HSCs undergo many distinct binary transitional steps, with defined hierarchical boundaries in order to produce mature blood cells. However, this model is mainly based on analysis conducted from FACS purified populations, in some cases followed by single cell analysis which rely on the lineage contribution associated with each population. The draw-back of these population based methods is that properties of the average population are investigated, whilst quantitative changes along with transitional states of cells between the pre-set gating strategies are not captured (Velten *et al.*, 2017). Several groups utilising RNA sequencing and computational reconstruction of pseudo-time, have revealed at the cellular level haematopoiesis is accomplished by a continuum of low-primed undifferentiated stem and progenitor cells that gradually differentiate into lineage restricted cells which occurs through a transcriptional program involving the repression of genes related to cell division and activation of lineage specific genes (Macaulay *et al.*, 2016; Velten *et al.*, 2017; Karamitros *et al.*, 2018). Another argument against these linear models is the growing evidence that the differentiation process of particular lineages is predetermined by HSCs early in haematopoiesis and that unipotent progenitors can differentiate directly from HSCs. This is supported by research observing that a subset of HSCs expressing the megakaryocyte-specific gene, Von Willebrand Factor (vWF) have robust short and long-term reconstitution ability post transplantation and that megakaryocyte differentiation can bypass MPP, CMP and MEP oligopotent stages (Månsson *et al.*, 2007; Notta *et al.*, 2016; Carrelha *et al.*, 2018). In support of this idea, a lineage-biased HSC population expressing high levels of Hdc (encoding histidine decarboxylase) has been documented, displaying myeloid-biased differentiation defined as myeloid biased (MB)-HSCs. (Chen *et al.*, no date; Oguro, Ding and Morrison, 2013). These populations highlight that HSCs can be defined into different lineage -biased subpopulations. At the molecular level they retain a common gene expression signature of HSCs though have

different chromatin architecture at lineage specific loci and exhibit lineage specific priming (Buenrostro *et al.*, 2018; Zhang *et al.*, 2018). Furthermore, *in situ* labelling and clonal tracking methods employed to assess clonal dynamics in adult mice at steady state, further challenging this dogma. These studies revealed that in an unchallenged setting, LT-HSCs have limited contribution to blood production throughout murine adulthood arguing a model in which recruitment of both lineage restricted and multipotent clones such as ST-HSCs and MPPs drive steady state haematopoiesis (Lu, 2014; Sun *et al.*, 2014).

1.1.2.2 Human hierarchy

In comparison to murine haematopoiesis, the study of human haematopoiesis has been limited due to the scarcity of available human tissues at early embryonic stages, along with the limited *in vitro* and *in vivo* models available to investigate purified HSCs (Doulatov, Notta, Laurenti and J. E. Dick, 2012). Whilst initial research in human haematopoiesis was limited to incidental morphologic observations, the first functional studies were achieved in 1960's where the regenerative potential of HSCs were assayed using *in vivo* repopulating assays- establishing the existence of multipotential HSCs (Becker, McCulloch and Till, 1963). *In vitro* studies have also progressed the field with the development of more specialised cultures supported by cytokine enrichment enabling the development of the Long term culture-initiating cell (LTC-IC) assay (Ramsfjell *et al.*, 1999). However, the most substantial development thus far is the use of successful engraftment of human haematopoietic cells in immunodeficient mice. These animals are capable of accepting human cells via transplantation, which engraft without rejection due to xenogeneic boundaries (Doulatov, Notta, Laurenti and J. E. Dick, 2012). Since their initial discovery various improved immunodeficient mice have been engineered, NOD/SCID/IL-2Rγnull (NSG), becoming routinely used in human studies (Brehm *et al.*, 2012). Through LTC-IC assays and transplantation experiments, human HSCs were first identified within the $CD34^+$ $CD38^-$ compartment (Larochelle *et al.*, 1996; Bhatia *et al.*, 1997) (Larochelle et al. 1996; Bhatia et al. 1997), which has been further refined to a $CD34^+CD38^-CD90^+CD45RA^-CD49f^+$ population (Wisniewski *et al.*, 2011). Consistent with the classical murine hierarchy model these HSCs ($Lin^-CD34^+CD38^-CD45RA^-CD90^+CD49f^+$) differentiate to multipotent progenitors ($Lin^-CD34^+CD38^-CD45RA^-$ $CD90^-CD49f^-$) that give rise to CMPs ($Lin^-CD34^+CD38^+CD45RA^-$ $CD135^+CD10^-CD7^-$) and multi-lymphoid progenitors (MLPs) ($Lin^-CD34^+CD38^-CD45RA^+$CD90-$CD135^+CD10^+$ $CD7^-$). Subsequently, MLPs have the potential to give rise to macrophage precursors and CLPs (Lin^- $CD34^+CD38^+CD45RA^+$ $CD10^+$) while CMPs bifurcate to GMPs (lin^-CD34^+ $CD38^+CD45RA^+CD135^+CD10^-CD7^-$) and MEPs ($Lin^-CD34^+$D38+$CD45RA^-$ $CD135^-CD10^-CD7^-$) oligopotent progenitors (Doulatov *et al.*, 2010; Notta *et al.*, 2011; Doulatov, Notta, Laurenti and J. E. Dick, 2012). This model was revised to include a CD34- HSC population ($CD34^-CD38^{low/-}CD45RA^-FLT3^-CD110^-CD133^+$GPI-$80^+$)

that sits at the apex of the hierarchy that has the capacity to differentiate along the MPP, CMPs GMP and MEP route whilst being capable of giving rise directly to MEPs (Sonoda, 2008; Sumide *et al.*, 2018) **(Figure 1.2)**.

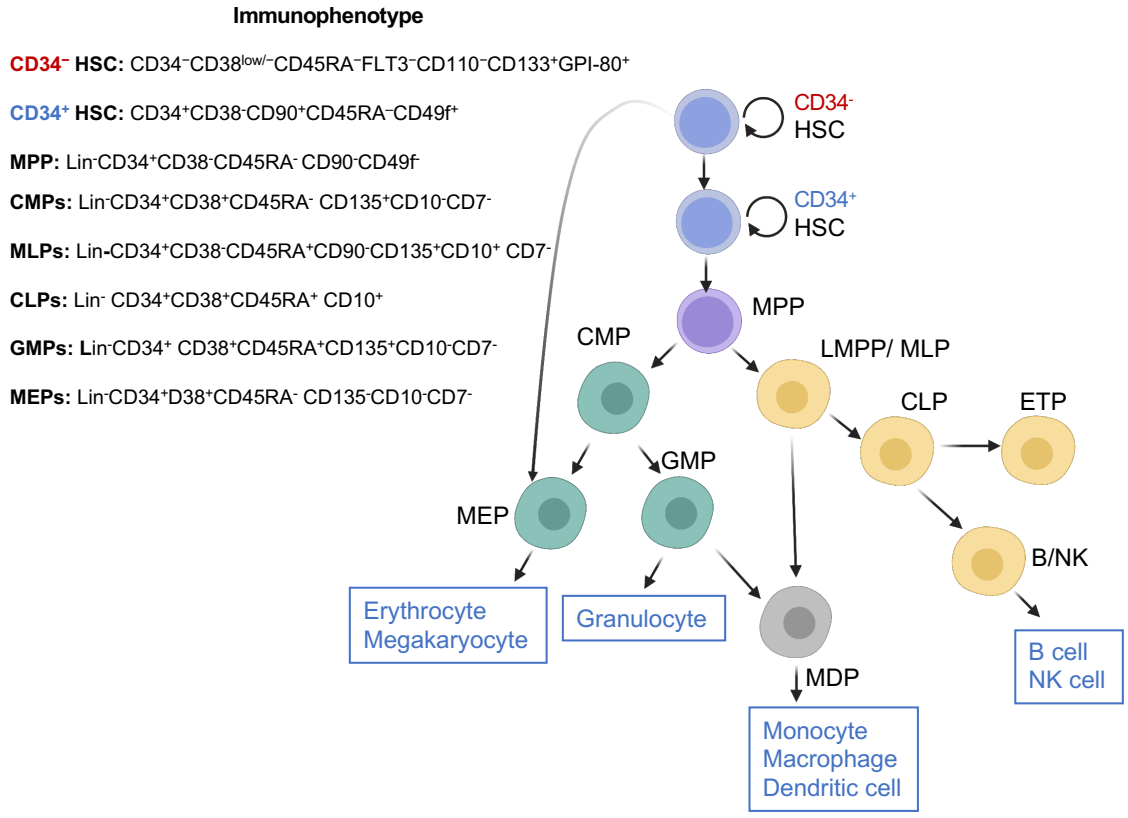


Figure 1.2. Human haematopoietic hierarchy. HSCs reside at the top of the haematopoietic hierarchy where they sustain lifetime haematopoiesis. HSCs give rise to multipotent progenitor cells with proliferative potential which also differentiate into the main blood cell types. CD markers that define each population are indicated.

1.1.3 Bone marrow niche

The BM niche was first conceptualised by Schofield in 1978 however, after decades of research our understanding of its localisation and function remains incomplete (Schofield, 1978). It is the primary site of haematopoiesis throughout adult life, comprised of a complex heterogeneous population of haematopoietic and non-haematopoietic cells surrounded by a shell of vascularized and innervated bone. It is the location where HSCs reside, are maintained and differentiate into multiple blood lineages. HSCs that lose contact with their associated niche cells undergo terminal differentiation, becoming committed progenitors and subsequently precursors of lymphoid cells, red blood cells (RBC), thrombocytes, granulocytes and monocytes. Cells undergoing differentiation are localised at the centre of the BM, where they proliferate and form colonies of maturing blood cells. Once fully matured these cells cross

the endothelium and enter into the blood stream. Moreover, immature lymphoid progenitors leave the BM and populate the thymus and lymphoid organs where they further differentiate (Calvi and Link, 2014; Morrison and Scadden, 2014). Most of the current knowledge of the BM niche has been revealed from murine studies of the long bones in mice. However, though all bones support haematopoiesis in mice, the major haematopoietic site in humans are the bones comprising the axial skeleton including the cranium, sternum, ribs, vertebrae and ilium (Yu and Scadden, 2016) **(Figure 1.3)**.

1.1.3.1 Regulation by non-haematopoietic niche cells

1.1.3.1.1 Osteolineage cells

Osteolineage cells were the first population of cells in the BM identified to regulate HSPCs even before Scofield's concept of the niche was suggested; these early studies reported that HSPCs were enriched in the endosteal region of the bone. Conversely, 3D imaging of the BM has revealed that HSCs are not significantly associated with osteoblasts in the niche (Kunisaki *et al.*, 2013; Nombela-Arrieta *et al.*, 2013). However, different osteolineage cells have been observed to play distinct roles in more committed progenitors of the lymphoid lineage. The importance of osteolineage cells in regulating B cells has been implicated by many studies mediated through their production of many regulatory molecules supporting each stage of differentiation from HSCs to IgM^+ B cells such as interleukin 7 (IL-7) and CXC-chemokine ligand 12 (CXCL12) (Tokoyoda *et al.*, 2004). Osteoblasts were shown to support the transition from $Rag2^-$ HSCs to $Rag2^+$ CLPs which was independent of other niche cells such as endothelial and adipocytes (Zhu *et al.*, 2007). Further work conducted by Yu *et al* (2016) revealed that early osteoblastic cells expressing Osterix (Osx^+) are responsible for producing abundant levels of IL-7 and insulin-like growth factor (IGF-1) molecules that regulate the maturation of committed B cells through the pro-B stage. This was identified in through deletion of Osx^+ cells or *Osx*-specific deletion of IGF-1 which led to B lymphopenia and loss of functional immunoglobulin production (Yu *et al.*, 2016). This is supported by previous work outlining that pre-pro-B cells, end stage B cells and plasma cells require CXCL12 and IL-7 for successful maturation (Tokoyoda *et al.*, 2004).

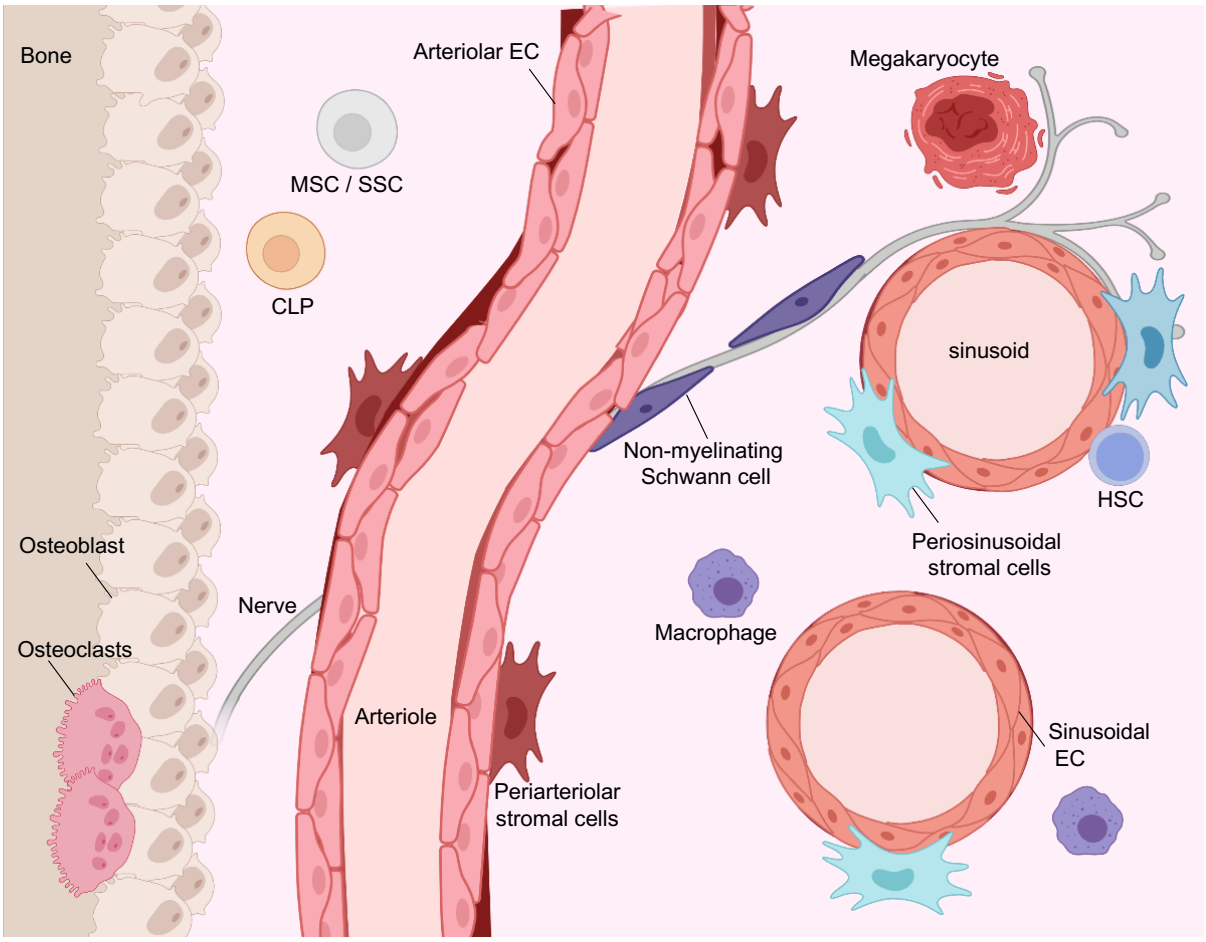


Figure 1.3. The bone marrow niche. HSCs reside close to the perivascular niches. The perisinusoidal niche is formed by fenestrated sinusoids, sinusoidal ECs, and perisinusoidal stromal cells (Lepr^{+}, CXCL12-abundant recitular CAR cells, NestinGFP^{dim}). The periarteriolar niche includes a central arteriole, arteriolar ECs, and periarteriolar stromal cells (NG2^{+}, NestinGFP^{high}). Other components of the BM niche include megakaryocytes, macrophages, nerve fibres and associated non-myelinating Schwann cells, MSC/SSC, and osteoblasts.

Additionally, target deletion of mature osteoblastic cells expressing Osteocalcin (Ocn^{+}) *in vivo* has been observed to result in decreased T-cell competent common lymphoid progenitors (CLPs) along with decreasing all stages of thymocytes from the earliest thymic progenitors to single positive CD4/CD8 expressing cells. A decrease in thymus homing receptor expression among BM haematopoietic cells was also observed due to the loss of Notch Ligand DDL4 activation on Ocn^{+} cells. This was further supported by selective depletion of DDL4 ligand from Oct^{+} cells resulting in thymopoietic defects. This work reveals that the BM osteolineage provides key molecular drivers enforcing thymus-seeding progenitor generation (Yu *et al.*, 2015). It has also been noted that the relative balance of Osx^{+} versus Ocn^{+} oesteoblastic cells in the BM are vital in creating a B versus T-progenitor niche therefore highlighting the

importance of these cells in the regulation of specific cell types of the adaptive immune system. Taken together, these studies highlight that lymphopoiesis is critically regulated by osteolineage, mesenchymal derived populations located in the BM niche. Mature osteolineage cells appear to be important during granulocyte colony-stimulating factor (G-CSF) mediate HSPC mobilisation, observed from studies depleting Osteocalcin (Ocn)-expressing cells which resulted in reduced G-CSF inducted mobilization (Ferraro *et al.*, 2011).

1.1.3.1.2 Perivascular cells

The observation that HSCs are frequently found to localise in close proximity to blood vessels rather than towards the endosteum has led to hypothesis that perivascular regions of the bone marrow may promote HSC maintenance (Kiel *et al.*, 2005; Pinho and Frenette, 2019). Mesenchymal stromal cells (MSCs) are rare, haematopoietic supporting cells that reside at the outer surface of sinusoids and wrap tightly around arterioles in the BM. They are primitive cells with self-renewal capacity, along with the ability to differentiate into osteolineage cells, chondrocytes and adipocytes (Yu and Scadden, 2016). These cells are an essential part of the niche and were originally described as a major source of CXCL12 which binds to CXC-chemokine receptor 4 (CXCR4) on HSCs, maintaining them within the niche. A range of markers have been used to define these cells in the BM such as CD146 and CD271 in humans along with CXCL12-GFP, Nestin-GFP, LEPR, Prx-1-Cre and Mx-1-Cre (Yu and Scadden, 2016).

CXCL12-abundant reticular (CAR) stromal cells were first discovered using a knock-in of a GFP reporter at the *Cxcl12* locus identifying them as adipo-osteogenic progenitors mainly distributed around sinusoids (Sugiyama *et al.*, 2006). They express high levels of CXCL12 and SCF- key factors involved in HSC maintenance (Sugiyama *et al.*, 2006; Omatsu *et al.*, 2010) along with IL-7 required for the maintenance of lymphoid progenitors and mature B cells (Cordeiro Gomes *et al.*, 2016). They are the major source of adipocytes and osteoblasts as observed from short term ablation of *in* vivo which led to severely impaired adipogenic and osteogenic differentiation of BM cells. Ablation of CAR cells has also been observed to led to a reduction in cytokine production (SCF and CXCL12) resulting in a marked reduction of cycling lymphoid and erythroid progenitors along with HSCs (Omatsu *et al.*, 2010) suggesting CXCL12-CXCR4 signalling is essential in maintaining a quiescent HSC pool (Sugiyama *et al.*, 2006). Transcription factor early B cell factor 3 (Ebf3) is preferentially expressed on these cells inhibiting CAR differentiation into osteoblasts whilst also maintaining CAR self-renewal. Studies employing conditional KO and retroviral systems demonstration that loss of Ebf3 expression in CAR cells impairs HSC abundance and leads to osteosclerosis. Showing Ebf3

expression is essential in regulating the interface between osteogenesis and haematopoiesis (Galán-Díez and Kousteni, 2018).

1.1.3.1.3 Endothelial cells

Endothelial cells are found lining the interior walls of all blood vessels (arterial and sinusoid) in the BM and are known to secret specific paracrine growth factors (angiocrine factors), cytokines and adhesion molecules such as; intracellular adhesion molecule 1 (ICAM), vascular cell adhesion molecule 1 (VCAM-1), E-selectin, and P-selectin in order to regulate homeostasis of HSPCs (Yu and Scadden, 2016). *In vivo* murine studies using endothelial-cell specific Cre recombinase strains have been used to selectively delete many of these factors from the BM such as *Cxcl12* which resulted in the reduction of HSC frequency (Ding and Morrison, 2013; Greenbaum *et al.*, 2013), gene encoding *gp130* cytokine receptor on EC resulting in BM hypocellularity and reduced HSC abundance (Yao *et al.*, 2005) and *Jag1* resulting in profound decreases in haematopoiesis and premature exhaustion of the adult HSC pool (Poulos *et al.*, 2013). These studies along with the use of human endothelial cells in maintaining *ex vivo* HSPCs in culture (Chute *et al.*, 2002, 2004) show the importance of endothelial cells in HSC regulation. HSPC maintenance in part is regulated by distinct blood vessel types with different permeability properties. Work conducted by Itkin *et al* (2016) has revealed less permeable arterial blood vessels maintain HSC in a low reactive oxygen-species (ROS) state, whilst more permeable sinusoids promote HSPC activation and are the sites for immature and mature leukocyte trafficking to and from the BM (Itkin *et al.*, 2016). Which also contradicts the state of hypoxia which has been long documented to be a fundamental feature of HSPCs independent of their localisation in the BM architecture (Nombela-Arrieta *et al.*, 2013).

Previously, the lack of markers available to separate EC subtypes; arterial endothelial cells (AECs) and sinusoidal endothelial cells (SECs) were lacking and therefore, the difference in the contribution to HSC maintenance was unclear. Recent research using a combination of Podoplanin (PCPN) and Sca-1 expression has enabled for the first time the separation of these populations in the bone marrow; AECs ($CD45^{-}Ter119^{-}$ $Sca\text{-}1^{bright}$ $PDPN^{-}$) and SEC ($CD45^{-}Ter119^{-}Sca\text{-}1^{dim}$ $PDPN^{+}$). Using this new isolation strategy it was revealed that AECs exclusively secrete SCF and that specific deletion of *Scf* in AEC and not SECs leads to significant reduction in HSC function (Xu *et al.*, 2018). Together this research suggests different HSC niches within the BM depending on their function. This idea is further supported by recent studies showing that lineage biased HSCs expressing vWF are associated with and are regulated by megakaryocytes whilst vWF^{-} HSCs were associated and regulated by arteriolar perivascular cells (Pinho *et al.*, 2018). However, this is controversial as others have

argued that quiescent HSCs are uniformly distributed throughout the BM (Acar *et al.*, 2015). Although the abundance of endothelial cells is comparable to the abundance of MSCs, research using a mouse model to knock in GFP to the *Scf* locus to track Scf expression revealed the level of SCF in EC are much less that what is observed in MSCs, which has also been demonstrated in CXCL12 expression (Ding *et al.*, 2012; Greenbaum *et al.*, 2013). ECs also play an important role post myelosuppressive injury by producing factors that promote haematopoietic regeneration such as jagged-2 (Kobayashi *et al.*, 2010; Doan *et al.*, 2013; Guo *et al.*, 2017).

1.1.3.1.4 Adipocytes

As previous stated adipocytes arise from MSCs and in adults represent 50-70% if the total BM volume. In humans the aging process accelerates BM adiposity which is especially apparent in the long bones where haematopoietic sites are replaced by fatty yellow marrow with reduced haematopoietic activity and is also associated with adipocyte-biased MSC differentiation (Kricun, 1985). Whether adipocytes are positive or negative regulators of HSC function is controversial. One of the earliest papers observed that during post sublethal irradiation, adipocyte abundance in the BM niche increased in parallel HSC proliferation. However, subsequent studies using pharmacological (peroxisome proliferator-activated receptor-γ-inhibitor bisphenol) or genetic (lipoatrophic A-ZIP/F1 'fatless' mice) approaches to reduce adipogenesis following irradiation of mice revealed an accelerated engraftment potential (Naveiras *et al.*, 2009). Additionally, reports observing HSCs are reduced in adipocyte-rich vertebrae of mice in comparison to adipocyte-lacking vertebrae of the thorax along with the adipocyte-secreted protein Adiponectin (ADIPOQ) being shown to impair proliferation and differentiation of HSPCs *in vitro* (Yokota *et al.*, 2000) all point to adipocytes being negative regulators within the BM niche.

In support of a more supportive role Zhou *et al* (2017) reported that though EC and $LepR^{+}$ stromal cells are critical sources of SCF in the BM, after their depletion during irradiation/ chemotherapy adipocytes increase in abundance and are responsible for generating SCF necessary for HSC reconstitution (Zhou *et al.*, 2017). This suggests a key role for adipocytes in promoting HSC factors for increased haematopoiesis in response to cytopenia. Importantly, it was noted that there are context-dependent differences in adipocyte function depending on BM site such as long bones in comparison to caudal vertebrae along with differenced between human and mice. This could account for the differences in positive versus negative effects observed in the discussed research. Furthermore, haematopoietic cell and adipocyte co-culture have shown a range of positive and negative effects on HSPCs. Adipocytes synthesise many factors such as adiponectin which promotes HSC proliferation and adipogenic signalling

is required for optimal HSC proliferation both *in vitro* and in long-term reconstitution *in vivo* (DiMascio *et al.*, 2007) along with leptin which promotes adipogenesis and inhibits osteogenesis by regulating MSCs in the BM (Poloni *et al.*, 2013; Yue *et al.*, 2016). It is possible that adipocytes secret a variety of factors that influence HSCs and stromal cells in the BM however, further studies are required to elucidate the function of adipocytes especially at steady state.

1.1.3.1.5 Macrophages

A subset of macrophages located in the BM niche have been observed to play a vital role in regulating HSC migration. It was found that $CD169^+$ macrophages secret oncostatin M (OSM) a soluble factor that induces CXCL12 expression possibly through a MEK-p38-STAT3 axis (Minehata *et al.*, 2006; Albiero *et al.*, 2015). As previously discussed *Nes*$^+$ MSCs and osteoblasts located in the niche are important sources of CXCL12 which is vital for HSC retention within the BM niche. This is further supported by the conditional depletion of $CD169^+$ macrophages by administration of clodronate-loaded liposomes into wild-type (WT) mice or macrophage Fas-induced apoptosis (Mafia) transgenic mice. These models observed reduced BM CXCL12 levels and selective downregulation of HSC retention genes in $Nestin^+$ niche cells, resulting in HSPCs exit into the blood stream (Chow *et al.*, 2011). It is possible that the crosstalk observed between $CD165^+$ and $Nestin^+$ MSCs in these experiments could be mediated through $CD165^+$ derived OSM. Furthermore, a rare population of BM monocytes and macrophages expressing high levels of α-smooth muscle (α-SMA) and the cyclooxygenase COX-2, have been identified adjacent to HSPCs within the niche. These cells under stress conditions, upregulate COX2 derived prostaglandin E_2 (PGE_2) protecting HSPCs from exhaustion by limiting the production of ROS though inhibition of the Akt kinase and stromal-cell expression of CXCL12 (Ludin *et al.*, 2012). These $CD165^+$ macrophages have been reported to also participate in supporting late erythroid lineage development through an unknown pathway. This was observed via specific deletion of $CD165^+$ macrophages in mice which led to a reduced abundance of erythroblasts, with no signs of anaemia in the BM steady state. Conversely, in a disease setting specific deletion was observed to cause haemolytic anaemia and acute blood loss due to impaired erythropoietic recovery (Yu and Scadden, 2016).

1.1.4 Regulatory transcription factors of haematopoiesis

Haematopoiesis is a complex process whereby multipotent blood stem cells give rise to a variety of phenotypically distinct mature cell types. Each cell within the hierarchy has their own distinctive gene expression pattern, controlled by various individual transcription factors (TF)

that are usually integrated into a wider regulatory network. It is the many combinations of lineage-determining transcription factors that drive; stem cell maintenance, lineage commitment from multipotent precursors and differentiation. Whilst TF are involved in activating gene expression, they often also have a repression mechanism which also plays an important role in lineage determination during haematopoiesis. Some unilineage-fated cells are derived from multilineage primed progenitors, which during differentiation need to undertake repressive mechanisms in order to become a specialised and mature cell. Moreover, genes that are required to enable cells to terminally maturate are primed during early stages of lineage commitment. Their expression also requires restraint to prevent premature, high-level expression which may leading to premature advanced differentiation (Porcher, Chagraoui and Kristiansen, 2017). Due to their high regulatory demand within the haematopoietic system, aberrations to these complex proteins often result in mis-expression, leading to haematological malignancies such as Leukaemia, highlighted extensively throughout the literature.

1.1.4.1 SCL/TAL

The transcription factor stem cell leukaemia (SCL), otherwise known as the T-cell acute lymphocytic leukaemia 1 (TAL1) contains a helix-loop-helix domain that binds DNA through its regulatory regions, interacting with the E-box sequence “CANNTG” along with GATA, Ets and Runx factor binding sites. SCL interacts with many protein partners in order to carry out activating and repressive transcriptional activities in a lineage and stage-specific manner. SCL’s activities are all mediated through nucleation of a core quaternary protein complex; SCL:E-protein:LIM domain-binding protein domain 1 or two (LMO1/2): LIM domain-binding protein (LDB1) and dynamic recruitment of other regulatory TF in a lineage and stage-specific manner (Porcher, Chagraoui and Kristiansen, 2017). SCL lies at the apex of a hierarchy of TF involved in haematopoietic specification due to its priming of *cis*-acting elements in gene loci involved in endothelial and blood specification (Porcher, Chagraoui and Kristiansen, 2017). SCL has two distinct roles during haematopoietic specification that have been described; in early developmental stages, low level expression and DNA binding are sufficient for lineage priming and gene repression during the specification process, conversely high levels and direct DNA binding are required for robust gene expression of primed genes during lineage maturation.

SCL is broadly expressed in adult blood cells though is absent in differentiating T and B lymphocytes. Nevertheless, SCL expression is shown to be activated in T-cells as a result of genetic aberrations such as translocations observed in 60% of childhood and adult T-ALL cases correlating with poor prognosis (Ferrando *et al.*, 2002, 2004). Full leukemic

transformation requires a cooperating genetic event such as LMO1/2 chromosomal translocations and activation of the NOTCH pathway (Ferrando *et al.*, 2002; Weng *et al.*, 2004; Gerby *et al.*, 2014). Moreover, though SCL expression is enriched in HSCs, its expression is not a requirement for HSC function at adulthood as observed in $Scl^{fl/fl}$ mice which had effective HSC engraftment, self-renewal and differentiation into myeloid and lymphoid lineages however, differentiation of into erythroid and megakaryocytic precursors was impaired. Taken together, this outlines SCL expression is essential for the genesis of HSCs though continuous expression is not required for HSC function (Mikkola *et al.*, 2003).

1.1.4.2 RUNX1

Runt-related proteins/ "Runx" are a family of three proteins containing a DNA-binding domain highly homologous to that of Drosophila runt. This domain interacts with a common heterodimeric partner CBF beta enabling Runx factors to bind to DNA effectively therefore regulating expression of their target genes (Wang *et al.*, 1996). Whilst initially Runx1 was found to be a critical regular of haematopoiesis, Runx2 for its role in bone development and Runx3 for its requirement in neurogenesis, it is now understood that all three members play roles in haematopoiesis. It is expressed in all emerging HPCs and HSC, therefore playing a vital role in the emergence of definitive haematopoiesis and the sustaining of normal stem cell function throughout adulthood (Bruijn and Dzierzak, 2017). RUNX1 also plays an important role in the repression of the erythroid gene expression program during megakaryocytic differentiation. During the bifurcation of megakaryocyte and erythroid commitment (MEP) transcription factors krueppel-like factor 1 (KLF1) and friend leukaemia intergration1 (FLI1) play instructive roles. Increased binding of RUNX1 to the locus of erythroid master regulatory KLF1 (along with the recruitment of co-repressors) during megakaryocytic differentiation counterbalances the activating role of TAL1 which results in increased expression of FLI1 (Kuvardina *et al.*, 2015). This shows RUNX1 is crucial in the regulation of megakaryocytic specification. Other studies have outlined key roles in megakaryocyte maturation and in thrombopoiesis (Goldfarb, 2009; Jalagadugula *et al.*, 2010; Draper *et al.*, 2017). Loss of function (LOF) germline mutations within *RUNX1* lead to familial platelet disorder with associated myeloid malignancies (FPDMM) whilst acquired LOF mutations and chromosomal rearrangements are frequently seen in patients with myelodysplastic syndrome (MDS) and many leukaemia's of both the myeloid and lymphoid lineages (AML, ALL and CML) (De Braekeleer *et al.*, 2011). Recent studies have shown that the growth and survival of leukaemia cells are dependent on wild type *RUNX1* expression in a variety of leukaemia's (Ben-Ami *et al.*, 2013).

1.1.4.3 C/EBPα

The leucine zipper, CCAAT-enhancer binding protein α (C/EBPα) is a well-studied lineage-specific TF involved in myeloid development. During cell fate decisions it primes and activates the myeloid expression programme by binding promoters and enhancers of myeloid related genes such as *CSF3R, IL-6R, CEBPE, GFI-1 or KLF5* (Avellino and Delwel, 2017) whilst competing with other TF (PU.1 and GATA2) to repress non-myeloid lineage genes in HSPCs to promote neutrophil differentiation (Avellino *et al.*, 2016). C/EBPα inhibits proliferation through induction of p21 (Timchenko *et al.*, 1996)(Wang *et al.*, 2003), has been shown to reduce apoptosis (Wang *et al.*, 2003)(Paz-Priel *et al.*, 2011) and is involved in maintaining stem cell quiescence (Hasemann *et al.*, 2014). Mutations in the open reading frame of C/EBPα has been observed in approximately 10% of AML patients resulting in truncated protein variants which prevent normal function and downregulation of myelopoiesis (Avellino and Delwel, 2017).

1.1.4.4 GATA Family

The GATA family of transcription factors is comprised of six members (GATA1-6) that modulate cell fate in various tissue systems, through the recognition of the nucleotide motif "(T/A)GATA(G/A)" in the promoter region of target genes (Merika and Orkin, 1993). Members have two zinc-finger (ZF) domains highly conserved between family members in which a zinc ion interacts with four cysteine residues, stabilising the ZF fold structure. Each of the ZF domains have a unique function with ZF1 (N-terminal) being responsible for binding the GATA recognition sequence, stabilizing the interaction with certain sequences or interaction with protein partners during protein complex formation, whilst ZF2 (C-terminal) is responsible for sequence recognition and critical DNA binding (Ko and Engel, 1993; Orkin, 1995; Harigae, 2006; Bates *et al.*, 2008). Unlike the two highly conserved ZF domains, the N and C terminal regions, composing transcription activation modules, diverge considerably between the individual GATA factors (Morrisey *et al.*, 1997). GATA family members were initially divided into two subgroups based on their tissue specific expression; haematopoietic factors (GATA1-3) (Ko and Engel, 1993)(Patient and McGhee, 2002) and cardiac factors (GATA 4-6). However, their function and expression patterns are not restricted to these tissues as observed in GATA2 and GATA3 which have important roles in skin, prostate, mammary gland, kidney and the central nervous system (Kaufman *et al.*, 2003; Kouros-Mehr *et al.*, 2006; Grote *et al.*, 2008). The activity of the GATA factors along with their role as transcriptional regulators are largely achieved through their interaction with co-regulators in order to assemble

transcriptional complexes and to recruit chromatin remodelling proteins (Tremblay, Sanchez-Ferras and Bouchard, 2018).

GATA1

The founding member of the GATA family, GATA1 is considered the master regulator of erythropoiesis due to its well documented gene regulatory roles during the maturation and differentiation of the erythroid lineage. Its role in the maturation of the erythroid lineage has been outlined from mouse knock-out studies where investigations showed that *Gata1* gene knockout results in embryonic lethality due to severe anaemia around day E10.5-11.5 along with deficient *GATA1* ES cells being able to contribute to all lineages, with exception of mature RBC during chimera studies (Pevny *et al.*, 1991). This was further supported by subsequent work showing the inability of *GATA1* null erythroid cells to mature into erythrocytes, arresting at the proerythroblast stage (Pevny *et al.*, 1995). *GATA1* participates in the transcription of most erythroid specific genes through its binding of GATA DNA motifs in gene regulatory regions (such as α and β-globin gene clusters) (Chang *et al.*, 2002). It also has the ability to down regulate cofactors necessary for granulocyte-monocyte and lymphoid development (*Spi1 (PU.1), Il7* and *Pax5*) whilst promoting megakaryocyte and erythrocyte commitment (Zaidan and Ottersbach, 2018a). Furthermore, in addition to erytho/megakaryocytic lineages, GATA1 is also expressed in eosinophils, basophils and mast cells. Whilst GATA1 expression is necessary for the development and function of eosinophils (Yu *et al.*, 2002) and basophils (Nei *et al.*, 2013), tissue specific knockouts models have indicated it is not necessary for the development of mast cells (Ohneda *et al.*, 2014). Taken together the tight control of *GATA1* levels are vital for many lineages of the haematopoietic system, this is further supported by the increased expression of *GATA1* documented in MDS patients (Hopfer et al 2012) along with *GATA1* mutations being associated with anaemia/thrombocytopenia, transient myeloproliferative disorders (TMD) and acute megakaryoblastic leukaemia (AMKL) in Down syndrome patients (Ahmed et al 2004).

GATA3

GATA3 has been shown to be a critical transcription factor in the development of several tissues and cell-lineages during developmental stages and in the adult. Within the haematopoietic system it has been extensively studied in innate and adaptive lymphoid development, where it is responsible for the regulation of differentiation and cell fate determination (Zaidan and Ottersbach, 2018). Initial experiments recognised it to be essential in the development, maintenance, survival and proliferation of early T-cell progenitors, as observed from the inability of murine ES-null *Gata3* to contribute thymocytic or T-cell lineages in chimera studies, along with identification of functionally important *GATA-3* binding sites in

multiple T-cell specific genes (Ting *et al.*, 1996). Moreover, *GATA3* has been shown to be required for both the induction and commitment of T-cell fate in early lymphoid progenitors, whilst also repressing the B lymphoid programme (Garćia-Ojeda *et al.*, 2013; Rothenberg, 2013). Within the T-cell lineage GATA3 is a master regulator of T helper type 2 cells (Th2), regulating differentiation by controlling genes encoding Th2 cytokines (Il4, Il5 and Il13) (Lee, Fields and Flavell, 2001). Throughout thymocyte development GATA3 is tightly balanced as increased expression has been shown to cause cytotoxicity (Nawijn *et al.*, 2001) whilst low expression results the cessation of thymocyte development due to lineage diversion (Taghon, Yui and Rothenberg, 2007). Outside of lymphoid lineages GATA3 is expressed at high levels in Long term-HSCs and plays a key role in balance of self-renewal and differentiation. (Buza-Vidas *et al.*, 2011; Frelin *et al.*, 2013). Aberrations in GATA3 expression has also been implicated in haematological malignancies; T-cell lymphoblastic leukaemia (T-ALL) (Sanda *et al.*, 2012), early T-cell precursor acute lymphoblastic leukaemia (ETP-ALL) (Zhang *et al.*, 2012) and Philadelphia chromosome-positive ALL (Ph-like ALL) (Perez-Andreu *et al.*, 2013).

1.1.5 GATA2

GATA2 is predominantly expressed within adult and developing haematopoietic stem cells (HSCs), myeloid progenitors and mast cells. It is regarded as a vital regulator of both HSC and haematopoietic progenitor cells (HSPCs). The murine *Gata2* gene is located on chromosome 6 and spans 8.5kb, whilst the human *GATA2* gene is located on chromosome 3 and spans 14kb. Both the murine and human gene encode five exons that result in two (murine) or three (human) mRNA transcripts, that each produce diverse isoforms (Rodrigues *et al.*, 2012). Both mouse and human *GATA2* genes generate proteins of between 466 and 480 amino acids of 49.1 kDa and 50.5 kDa. Two of the three mRNA transcripts have been studied in-depth and contain two untranslated first exons, IS and IG which are conserved among vertebrates and are differentially expressed. The longest, GATA2-IS shows activity specific to haematopoietic and neuron cells, whilst the shorter form (IG) is active in all tissue types expressing *GATA2* (Minegishi *et al.*, 1998; Pan *et al.*, 2000) **(Figure 1.4 A)**.

Five highly conserved cis-acting sites are involved in the regulation of *GATA2* expression; -77kb, -3.9 kb, -2.8, -1.8kb and +9.5 named relative to their location from the IS promoter. All sites contain GATA DNA motifs allowing GATA2 itself to bind at each site implying positive autoregulation (Soukup and Bresnick, 2020). In parallel, other activators or repressors are able to displace and modulate Gata2 expression such as GATA1-GATA2 switching, which occurs at each of the five cis-regulatory elements **(Figure 1.4 C)**. Knock out mouse models lacking each of these individual sites have revealed variable effects to Gata2 expression and haematopoiesis **(Table 1.1).** The -2.8 site has been shown to be responsible for maintaining

high levels of Gata2 expression required in HSPCs (Kirby D Johnson *et al.*, 2012; Gao *et al.*, 2013), whilst for efficient terminal differentiation of erythroblasts, Gata1 competes for occupancy at the -1.8 site resulting in displacement of Gata2 (Moriguchi and Yamamoto, 2014). The role of -3.9 site has yet to be defined as mice lacking this enhancer revealed no alterations to Gata2 expression or haematopoiesis (Katsumura *et al.*, 2017). Though these sites (-1.8 and -2.8) play enhancer roles during GATA2 transcription, their deletion has revealed they are not essential for Gata2 expression and haematopoiesis (Snow *et al.*, 2010).

On the other hand, the -77 and +9.5 site are known to have essential functions in embryogenesis and haematopoiesis with deletions in mouse models causing embryonic lethality at days E15.5 and E13-14 respectively. The -77 site was shown to be involved in a crucial part of the myeloid progenitor transcriptome, upregulating Gata2 expression in myeloid progenitors, with its deletion impairing differentiation (Johnson *et al.*, 2015). Further investigations highlighted in particular, the -77 site regulated MEPs as reduction in erythroid progenitors were observed in compound mouse models due to the inability to induce Gata1 expression (Mehta *et al.*, 2017). The +9.5 site enhances *Gata2* transcription during many developmental stages of embryogenesis where it plays a vital role in the generation of HSC in the AGM along with various other tissues (Kirby D Johnson *et al.*, 2012) and in controlling GATA2 levels in HSPCs and erythroid precursors (Soukup and Bresnick, 2020).

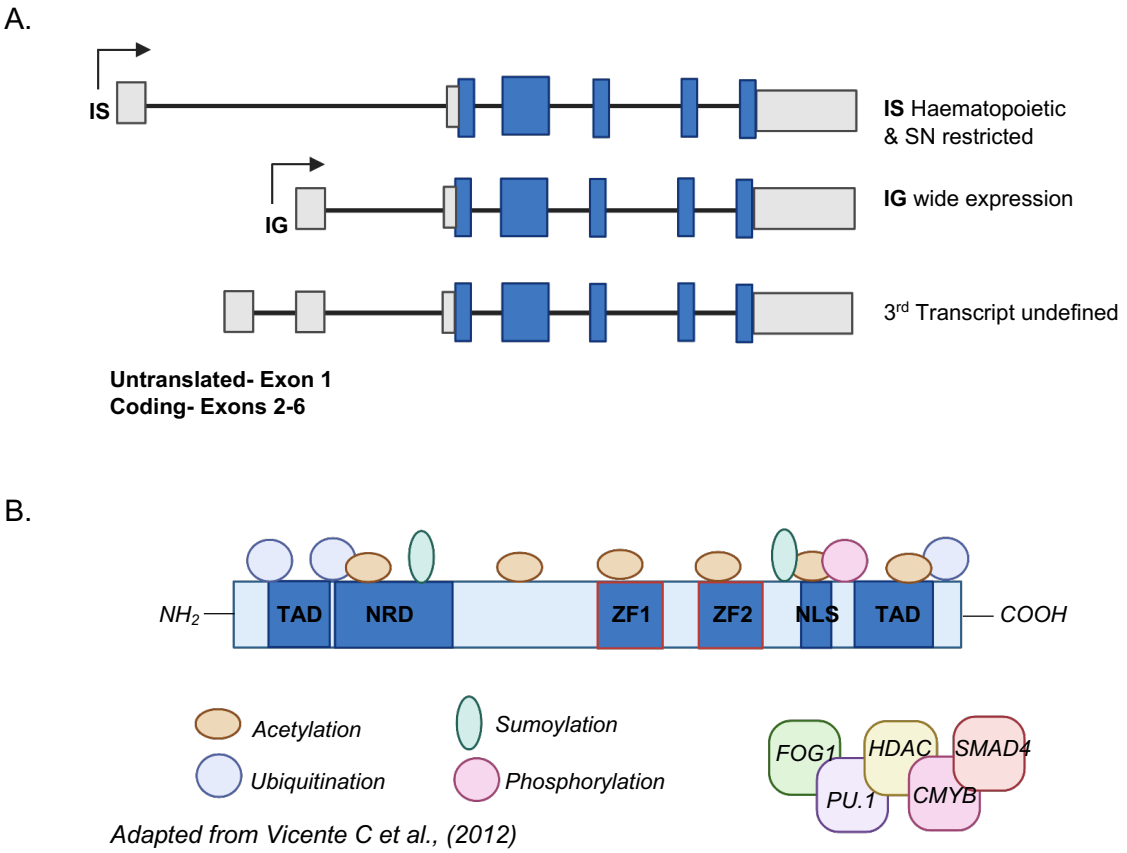


Figure 1.4. *GATA2* isoform and protein regulation. (A) Human *GATA2* gene structure showing IS, IG and a 3rd transcript that has yet to be defined. (B) *GATA2* protein structure outlining post-translational modifications (PTM) including ubiquitination, acetylation, sumoylation, and phosphorylation and protein-protein interactions with *FOG1, PU.1, HDAC, CMYB at either zinc finger 1 or 2 domains*. (C) *GATA2* cis regulatory elements termed according to their relative position to IS promoter: -77 kb, -3.9 kb, -2.8, -1.8, +9.5.

The protein is composed of two zinc finger DNA binding domains. It is known that the function of GATA-2 in modulating haematopoietic cells fate is dependent on the two conserved zinc fingers (N and C fingers) and their flanking sequences, including two transactivating domains, a nuclear localization signal and a negative regulatory element (Minegishi *et al.*, 2003).

Table 1.1 GATA2 Cis-acting *sites and function*

Cis-acting site (kb)	(-/-) KO viability	(-/-) KO Gata2 expression	Effect on haematopoiesis	Regulation of Gata2	REF
-77	N (E15.5)	Reduced (GMP & MEP)	Normal HSC levels Reduction of Myeloid progenitors Reduced erythroid progenitors	**Enhancer** Essential for embryogenesis Essential for MEP progenitor cell multipotentiality Myeloid & erythroid progenitor differentiation	(Johnson *et al.*, 2015)
-3.9	Y	ND	NE	**Enhancer** Unknown function	(Sanalkumar *et al.*, 2014)
-2.8	Y	Modest reduction (HPSCs)	Late E lethality Anaemia	**Enhancer** Maintaining high expression in HSPC	(Sanalkumar *et al.*, 2014)
-1.8	Y	ND	Erythroid maturation block (Gata1 dysregulation) -loss Gata1 occupancy	**Repression** Essential for terminal differentiation erythropoiesis	(Snow *et al.*, 2010)
+9.5 'master regulatory *cis*-element'	N (E13-14)	Reduced	Abolished HSC generation (AGM)	**Enhancer** Essential for embryogenesis, HSCs (AGM) Essential for HSPC and erythroid precursors Establishment of definitive FL HSPC compartment (not for primitive) SC generation	(Kirby D. Johnson *et al.*, 2012; Gao *et al.*, 2013)

1.1.5.1 GATA2 function in haematopoiesis

GATA2 expression and function differs in stem and progenitor compartments which could be attributed to the different cell types that GATA2 is expressed in. Stem and progenitor pools express distinct TF and therefore GATA2 may interact with different protein partners in each compartment. It is now well-established that GATA2 levels require tight regulation at all developmental stages for normal haematopoietic function. Numerous studies have explored the role of GATA2 in HSC and HPCs using both loss and gain of function models. Germline deficient *Gata2*$^{-/-}$ mice die during gestation at embryonic day 10-10.5 due to anaemia, with vast reductions in both primitive and definitive HPCs in the YS. Further analysis using chimeric embryo generation with *GATA2*$^{-/-}$ ES cells and *in vitro* differentiation models have outlined GATA2 to be essential in the expansion of multipotent haematopoietic progenitors and in the formation of mast cells, though is dispensable for the terminal differentiation of erythroid and macrophage cell types (Tsai *et al.*, 1994; Tsai and Orkin, 1997).

1.1.5.1.1 Role in HSCs

1.1.5.1.1.1 HSC generation and function

During embryonic development HSC are generated from hemogenic endothelium in the ventral wall of the AGM region (Dzierzak and Speck, 2008; Ivanovs *et al.*, 2014) and their specification is tightly controlled by temporal changes in master regulators. GATA2 was first identified as a key component in HSC development when studies observed the collapse of primitive and definitive haematopoiesis in *GATA2* knock out mice leading to embryonic lethality at E10.5 (Tsai *et al.*, 1994; K.-W. W. Ling *et al.*, 2004). Further work by Tsai and Orkin (1997), demonstrated that *Gata2*-deficient stem cells and yolk sac cells proliferate poorly and undergo extensive necrosis outlining the requirement for GATA2 for the expansion of multipotent haematopoietic progenitors. Additional research has investigated the role of GATA2 through GATA2 haploinsufficiency, which revealed that *Gata2*$^{+/-}$ mice showed a marked reduction of early progenitors and severe reduction in the numbers of HEC, IAHCs, HPCs and HSCs, along with qualitative defects in HSC as shown from their poor performance in competitive transplantation experiments (K.-W. W. Ling *et al.*, 2004; Rodrigues *et al.*, 2005; Gao *et al.*, 2013). Similarly, conditional knockout of *GATA2* in VE-cadherin (VEC) endothelial cells in conjunction with analysis of haematopoiesis from the AGM region in mice with deleted *Gata2* +9.5 cis-element revealed its requirement for the formation of intra-aortic haematopoietic clusters and HSCs (Lim *et al.*, 2012; de Pater *et al.*, 2013). Together these studies highlight a regulatory effect of GATA2 during the HE and EHT stages of development, though a specific mechanism was unknown. More recently a similar role of GATA2 promoting EHT has been observed to translate in the human setting. Key work published by Kang *et al* (2018) utilising a hESC line with conditional GATA2 expression (iG2$^{-/-}$ hESC) was used to

examine the effect of GATA2 at specific stages of haematopoiesis. The authors revealed that HE development and diversification into arterial and non-arterial HE was regulated independently of GATA2, thought was vital for the transition of HE to EHT (Kang *et al.*, 2018). In support of this notion the reprogramming of human fibroblasts into hemogenic cells expressing dynamic EHT transcriptional programs has been achieved using ectopic expression of *GFI1B, FOS* and *GATA2*. *GATA2* was revealed as the dominant TF involved in the activation of the hemogenic program at early reprogramming stages (Zhou *et al.*, 2019). Using iGATA2-hiPSCs GATA2 expression at the mesoderm specification stage promotes the generation of hemogenic endothelial progenitors and there subsequent differentiation to HPC whilst acting as repressors at cardiac gene promotor regions (Castaño *et al.*, 2019). Additionally, a study employing the GATA2Venus reporter mouse model, with unperturbed GATA2 expression, observed that whilst all HSC are GATA2 expressing, not all HPC in the aorta, vitelline and umbilical arteries and FL require or express GATA2. Therefore GATA2 is vital during the programming towards a HSC fate, whilst also corroborating that its expression is not required in all cells undertaking a HPC fate (Kaimakis *et al.*, 2016).

1.1.5.1.1.2 HSCs in adult BM

In adult life haematopoiesis is dependent on HSC and progenitors regulated by a combination of cell extrinsic factors instructed by the BM niche and a cell-autonomous programme determined by TF responsible for the regulation of gene expression. Though the transcriptional programs involved in differentiation stages in myeloid and lymphoid commitment have been characterised extensively using mouse models (Orkin, 2000), only a handful of TF have been outlined to be important in the transcriptional programmes involved in adult HSC survival thus far (Hock *et al.*, 2004; Kranc *et al.*, 2009; Pajerowski *et al.*, 2010). Mouse transplantation studies observe that enforced expression of GATA2 blocks the amplification and differentiation of HSC post transplantation. Therefore, down-regulation of GATA2 during haematopoietic lineage commitment is a necessity for *in vivo* haematopoiesis (Persons and Allay, 1999).

Due to the lethality observed during *GATA2* deletion, *Gata2* haploinsufficent mice expressing reduced Gata2 have also been used to assess the effect of *Gata2* gene dosage in adult haematopoiesis. It was observed that BM of $Gata2^{+/-}$ mice showed reduced amount of HSC along with perturbed functionality assessed by the lack of haematopoietic reconstitution during transplantation assays (Rodrigues *et al.*, 2005). Further investigation revealed this effect was associated with a decrease in HSC proliferation and an enhanced HSC quiescence and apoptosis, with an associated downregulation of anti-apoptotic and *GATA2* binding target Bcl-xL (K.-W. W. Ling *et al.*, 2004; Rodrigues *et al.*, 2005). Investigations into the genetic

requirement for *Gata2* in adult haematopoiesis either by ER-Cre mediated deletion, in the Gata2 C-terminal zinc finger domain or conditional deletion by the Mx1-Cre system, in line with these findings, showed an acute loss of adult HSCs and multilineage potential. Moreover, reducing GATA2 expression in human $CD34^{+}$ cells profoundly altered HSPC function as observed from CFC and LTC-IC assays mirroring murine studies and therefore identifying GATA2 as a key regulator of human HSPCs (H. S. Li *et al.*, 2016; Menendez-Gonzalez *et al.*, 2019).

Various experiments employing different means of enforced expression of GATA2; tamoxifen, retroviral or doxycycline (DOX) based approaches, also observed a reduction in proliferation, a block in differentiation and induced quiescence in both mouse and human studies (Persons and Allay, 1999; Kitajima *et al.*, 2002; Tipping *et al.*, 2009). Both murine and human HSPC expressing *Gata2/GATA2* displayed reduced performance during CFC and LTC-IC assays, whilst once again presenting with an impaired contribution to human haematopoiesis in *vivo.* Transcriptional profiling has shown GATA2 to be expressed highly in quiescent haematopoietic cells (Ezoe *et al.*, 2002; Venezia *et al.*, 2004) and known HSPC cell cycle regulators (p21 and p27) and cyclin-dependent kinases (Cdk4 and Cdk6) have been shown to be modulated by GATA2 (Kitajima *et al.*, 2002; Tipping *et al.*, 2009). Furthermore, studies have demonstrated phosphorylation of GATA2 at (Cdk) consensus motifs and report the interaction of GATA2 with Cdk2 and Cdk4 phosphorylated GATA2 *in vitro.* This suggests that GATA2 phosphorylation by Cdk/cyclin systems could be responsible for the cell cycle dependent regulation of GATA2 expression (Koga *et al.*, 2007). Taken together, these studies have outlined GATA2 as a key TF, indispensable for adult HSC survival and maintenance through cell cycle regulation in a critical level dependent manner.

1.1.5.1.2 Role in myelo/erythroid progenitors

1.1.5.1.2.1 Myeloid progenitors (GMPs)

Though GATA2 expression in myeloid cells is significantly lower than in HSCs, BM from GATA2 heterozygote mice revealed perturbed granulocyte-macrophage progenitor function in CFC assays, which was subsequently identified to be from the immunophenotypically defined GMP compartment. Additional analysis showed not only a reduction in the population size but impaired functionality, documented during competitive transplantation assays (Rodrigues, Boyd, Fugazza, May, YanPing Ping Guo, *et al.*, 2008). Attenuated expression of Notch target, Hes-1 was observed from $GATA2^{(+/-)}$ GMP cells. Binding of GATA2 to the *HES-1* locus was confirmed in myeloid progenitor cell line 32Dcl3 and overexpression of Hes-1 in heterozygous GMP cells was observed to recover the $GATA2^{(+/-)}$ phenotype. (Rodrigues, Boyd, Fugazza, May, YanPing Ping Guo, *et al.*, 2008). Interestingly, moderate overexpression of GATA2 using

an inducible GATA2-estrogen receptor (ERT) vector revealed enhanced self-renewal of myeloid progenitors *in vitro* resulting in immortalisation of BM cells to produce myeloid cells lines. Furthermore, *in vivo* transplantation assays showed an expansion of myeloid cells at the GMP stage along with block in B and T cell differentiation at the CLP population (Nandakumar *et al.*, 2015). This suggests that even with the low levels in myeloid cells *Gata2* acts as a critical regulator in GMP function.

1.1.5.1.2.2 Megakaryocyte/erythroid progenitors (MEPs)

The lineage and stage-specific gene expression during erythroid differentiation is coordinated by "GATA switching" that occurs between GATA1 and GATA2. As previously mentioned GATA2 is predominantly expressed in HSCs whilst GATA1 is undetectable. During erythroid differentiation, *GATA1* gene regulatory regions become demethylated allowing abundant GATA2 protein to bind to its GATA sites promoting GATA1 gene expression in early MEPs (Takai *et al.*, 2013). GATA1 subsequently regulates the expression of a sequence of erythroid-specific genes to allow further erythroid commitment where at later stages of differentiation its expression is further upregulated through an autoregulatory mechanism. During this process the abundant level of GATA1 protein interacts with Fog1 and other co-repressors displacing GATA2 at the GATA1 binding sites, whilst the GATA1 protein also supresses GATA2 expression by replacing GATA2 at the binding site of the *GATA2* gene, allowing terminal differentiation (Suzuki, Shimizu and Yamamoto, 2011; Suzuki *et al.*, 2013). Studies in mouse strains lacking the endogenous-GATA cis acting sites **(Table 1)** demonstrate the vital impact of these regions in this "GATA switching" outlined above. Genetic deletion of the -1.8kb cis acting site revealed erythroid maturation block due to a loss of GATA1 occupancy, along with the -77 kb site being detrimental in MEP regulation outlining the importance of GATA1-GATA2 binding at these regulatory elements during erythropoiesis (Fujiwara *et al.*, 2009; Kirby D. Johnson *et al.*, 2012; Mehta *et al.*, 2017).

1.1.5.1.3 Role in Mature cells

1.1.5.1.3.1 Mast cell, eosinophil and basophil function

GATA2 has been shown to be critical in both basophils and mast cell differentiation (Tsai and Orkin, 1997; Iwasaki *et al.*, 2006). The order in which GATA2 and C/EBPα is expressed has been shown to be vital in determining basophil fate, where if GATA2 expression is proceeded by CEBPα expression at the GMP stage basophil differentiation is induced, where as if the role is reversed eosinophil differentiation is driven (Iwasaki *et al.*, 2006). GATA2 has also been demonstrated as essential for the differentiation of pre-BMPs (FcεRIα+ GMPs) into basophils and mast cells, along with the maintenance of cell identity by regulating FcεRIα expression in basophils (impacting cell survival) and FcεRIα and c-Kit expression in mast cells (Li *et al.*,

2015). Moreover, GATA2 has also been implicated as a positive regulator of FcεRIβ, as a direct activator of the *Ms4a2* promoter (Ohneda *et al.*, 2014; Ohmori *et al.*, 2015, 2019). Furthermore, GATA2 was shown to be crucial for maintaining genes encoding cytokines Il4 (basophils) and Il3 (mast cells) that are important for performing basophil or mast cell functions and in the synthesis of histamine by upregulating Hdc expression (Li *et al.*, 2015). Taken together, it is clear that GATA2 plays essential roles in mast cells, eosinophils and basophil differentiation, function and cell identity.

1.1.5.1.3.2 Dendritic cells (DCs)

DCs play a vital role in the immune system and have been observed to be reduced in a number of immunodeficiency syndromes, including GATA2 deficiency which results from germline haploinsufficient mutations in the *GATA2* gene, of which DCs are profoundly decreased. As a result of this correlation many groups have looked to elucidate the role of GATA2 in DC differentiation. Gata2 mRNA levels have been reported at low levels in CMPs, detectible in GMPs though are undetectable in steady-state splenic cDCs ($B220^{-}CD11c^{hi}$) and pDCs ($B220^{+}CD11c^{lo}$), in contrast to *in vitro* derived FLT3-L–induced DCs and GM-CSF–induced DCs which are detectable (Onodera *et al.*, 2016). In spite of the little to no expression of GATA2 in DCs, *Gata2* deletion *in vivo* reveals profound decreases in DC abundance. Investigation into the ability of each progenitor compartment to give rise to DCs revealed LSKs, CMPs, common dendritic progenitors (CDPs) but not CLPs or GMPs to have an impaired DC generation, suggesting a role for GATA2 in myeloid DC differentiation. Furthermore, expression profiling from *Gata2*-knock out DC progenitors revealed the downregulation of several myeloid-related genes and surprisingly the upregulation of major T-cell- related genes such as *Tcf7* and *Gata3*. Therefore, GATA2 has an important cell-fate specification role towards the myeloid lineage whilst suppressing T-lymphocyte lineage differentiation by regulating lineage specific TF in DC progenitor genes therefore contributing to DC differentiation (Onodera *et al.*, 2016).

1.2 Myeloid malignancies

1.2.1 Myelodysplastic syndrome (MDS)

Myelodysplastic syndromes (MDS) are a heterogenous group of clonal myeloid neoplasms characterised by progressive cytopenia in the peripheral blood due to defective haematopoiesis, morphological features of dysplasia in the bone marrow and a strong propensity for clonal evolution and progression to AML (Steensma, 2018c). MDS can occur at any age however, it is most commonly observed in older individuals over the age of 70 and approximately 15-20,000 new cases are reported each year. Patients often present with

progressive macrocytic anaemia followed by pancytopenia. Diagnosis is often challenging due to the diversity of conditions across the spectrum of MDS and the similarity of diagnostic parameters with other related disorders such as MDS/ myeloproliferative neoplasms (MPN) disorder chronic myelomonocytic leukaemia (CMML) (Tanaka and Bejar, 2019). Genetic defects such as aberrant cytogenetics, gene mutations and altered epigenetic regulation play a major role in the pathogenesis of MDS.

1.2.1.1 Molecular pathogenesis

Clonal cytogenetic abnormalities have been detected in around 50-60% of MDS patients comprising; deletions of the long arm of chromosome 5q and 7q along with trisomy 8 and complex karyotypes **(Table 1.2)**. These are usually found at disease presentation and have been shown to play an important role in disease progression.

Table 1.2 MDS chromosomal abnormalities and their frequencies at diagnosis (Hosono, 2019)

Chromosomal abnormality	MDS (%)	Therapy related MDS (%)
Loss of chromosome 7/ del(7q)	10	50
Del(5q)	10	40
Gain chromosome 8	10	-
Del (20q)	5-8	-
Loss of Y chromosome	5	-
Isochromosome 17q/ t(17p)	3-5	25-30
Loss of chromosome 13/ del (13q)	3	-
Del(11q)	3	-
Del(12p)/ t(12p)	3	-

Interstitial deletion of the long arm of chromosome 5 (del(5q)) is the most abundant chromosome defect observed in 10-15% of MDS patients and is also observed in up to 40% of patients with secondary AML (Hosono, 2019). The most commonly deleted region (CDR) is a distal break point at band q32-q33.1 containing genes implicated in haematopoiesis and is mostly observed in patients with 5q syndrome, which has a low rate of AML progression. Functional studies have revealed individual genes that contribute to the clinical phenotype through haploinsufficient gene expression. Heterozygous loss of *RPS14* has been linked to defective erythropoiesis through activation of the p53 pathway, along with perturbed expression of miR-145 and 146a which is involved in innate immune signalling in HSPCs (Ebert *et al.*, 2008; Starczynowski *et al.*, 2010). Furthermore, dysregulation of Casein kinase I isoform alpha (*CSNK1A1)* a tumour suppressor gene located at 5q32 has been observed to activate b-catenin inducing HSC expansion (Schneider *et al.*, 2014).

Deletion of the extreme telomeric regions (5q34 to the telomere) is associated with an aggressive disease outcome. In these cases, DEAD-Box Helicase 41 (*DDX41*) expression (located at 5q35.3) is significantly decreased. *DDX41* haploinsufficiency is a known predictor of poor survival as observed from familial MDS/AML syndromes characterised by an inherited DDX41 mutation (Polprasert *et al.*, 2015; Maciejewski *et al.*, 2017). This suggest that the loss of function of DDX41 pays a pathogenetic role in MDS disease progression in these cases. The majority of myeloid neoplasms with del(5q) events are morphologically heterogeneous, are associated with a poorer prognosis and are commonly observed with other cytogenetic aberrations. Taken together, this suggests that this region of chromosome 5 has a key pathogenic role in disease phenotype and clonal evolution of myeloid malignancies.

Monosomy 7 and del(7q) occur recurrently throughout myeloid neoplasm accounting for; 10% *de novo* MDS, 50% of therapy related MDS, observed in MDS/myeloproliferative neoplasms (MPN) and primary and secondary AML and associated with poor prognosis (Bernasconi *et al.*, 2005; Haase *et al.*, 2007; Hosono, 2019). Studies employing SNP-array-based karyotyping methods along with metaphase cytogenetics have identified CDR (Jerez *et al.*, 2012; Hosono *et al.*, 2014) encoding vital genes such as; Cut Like Homeobox 1 (*CUX1)* which act as a tumour suppressor in myeloid progenitor cells regulating genes involved in cell cycle and DNA repair (McNerney *et al.*, 2013; Papaemmanuil *et al.*, 2013), LUC7 Like 2 (*LUC7L2)* a mammalian splicing factor involved in myeloid differentiation (Hershberger *et al.*, 2016), Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (*EZH2)* a member of the polycomb repressive complex 2 (PRC2) associate with trimethylation of H3K27 (Hosono *et al.*, 2014) and Sterile Alpha Motif Domain Containing 9 Like *(SAMD9L)* which negatively regulates cell proliferation acting as a tumour suppressor gene in haematopoietic cells (Nagamachi *et al.*, 2013; Inaba, Honda and Matsui, 2018). Isolated trisomy 8 is observed in 5-7% of MDS cases, is classified as an intermediate cytogenetic risk group and has been observed as a secondary or late event during MDS transformation (Paulsson *et al.*, 2001). How this genetic event contributes to MDS disease remains uncertain, however studies have revealed CD34 cells from MDS patients with trisomy 8 show resistance to apoptosis through upregulation of downstream anti-apoptotic proteins c-myc, Surviving and Cluster of differentiation 1 (CD1) (Sloand *et al.*, 2007).

Recurrent somatic gene mutations have been observed 80-90% of MDS patients, in more than 50 genes. Somatic mutations involved in spliceosome machinery have been reported in up to 50-89% of MDS patients, are heterozygous, occure in an exclusive manner and are associated with disease development (Yoshida *et al.*, 2011). The spliceosome is composed

of a core of small nuclear ribonucleoproteins (snRNPs) vital to the structural and enzymatic function and accessory proteins (non-snRNPs) that are important for structural assembly, splice site selection and co-ordination of alternative splicing events (Lindsley and Ebert, 2013). Its function is to mediate intron excision followed by ligation of exons to generate mature messenger RNA. Therefore, mutations in splicing factors result in differential splicing patterns that dysregulate various target genes which drive disease. The exact molecular consequences however, are yet to be fully defined (Pellagatti *et al.*, 2018; Taylor and Lee, 2019). Commonly mutated spliceosome genes include Splicing Factor 3b Subunit 1 (*SF3B1*) (20-25%), Serine And Arginine Rich Splicing Factor 2 (*SRSF2*) (10%) and U2 Small Nuclear RNA Auxiliary Factor 1 (*U2AF1*) (8%). Furthermore, *DDX41* and *LUC7L2* are also part of the splicing pathway and mutations in these genes have also been reported in MDS patients (Quesada *et al.*, 2019). Along with genetic lesions in RNA splicing, mutations in genes encoding epigenetic modifiers have also been implicated in MDS and are associated with widespread alterations in transcriptional programs (Lindsley and Ebert, 2013). Mutations involved in DNA methylation such as DNA Methyltransferase 3 Alpha (*DNMT3A*) and Tet Methylcytosine Dioxygenase 2 (*TET2*) are also some of the most commonly reported in MDS patients result in widespread aberrant promoter hypermethylation leading to MDS pathogenesis (Figueroa *et al.*, 2009). *DNMT3A* mutations are particularly associated with a poor prognosis and an increased risk of progression to AML (Hosono, 2019). IDH1/2 mutations have been observed in 5-10% of MDS and AML patients. Other mutations in histone modifiers (*ASXL1* and *EZH2*) and heterozygous deletions resulting in haploinsufficient gene expression such as *DDX41, LUC7L2, EZH2* and *CUX1* have also been implicated in MDS (Lindsley and Ebert, 2013; Papaemmanuil *et al.*, 2013; Kennedy and Ebert, 2017).

1.2.1.2 Diagnosis and treatment

Prior to treatment patients are risk-stratified to aid with treatment regimens according to the 2012 revised International Prognostic Scoring System (IPSS). This system takes into consideration the number and degree of cytopenias, percentage of blast cells present and risk of cytogenetic abnormalities, which places patients in one of five risk groups; two low risk, one intermediate and two high risk groups (Greenberg *et al.*, 2012). Patients diagnosed with mutations in Tumor Protein P53 (*TP53),* Neuroblastoma RAS Viral Oncogene Homolog *(NRAS),* Additional Sex Combs Like 1 *(ASXL1)* and *EZH2* are associated with poor outcomes in comparison to those with *SF3B1* mutations which is associated with more favourable outcomes (Bejar *et al.*, 2012; Papaemmanuil *et al.*, 2013). Patients with lower risk MDS are observed with mild cytopenias or present asymptomatically at diagnosis. These patients are monitored until cytopenias progress or symptoms present as early treatment has not shown to be beneficial in preventing clonal evolution or mortality. When patients exhibit severe

cytopenias, treatment is aimed to alleviate symptoms, preventing complications such as bleeding and severe infections and improving quality of life (Table 1.3). To alleviate the anaemia in these cases 'off the label' drugs such as erythropoiesis stimulating agents (ESA) epoetin and darbepoetin are used and in cases of genetically defined del(5q) MDS lenalidomide is administered (Talati, Sallman and List, 2017; Platzbecker, 2019). Depending on the molecular pathogenesis, some low-risk patients are also candidate for immunosuppressive therapy, thrombopoiesis-stimulating agents or DNA hypomethylating agents (HMA) such as Azactidine and Decitabine.

Allogeneic haematopoietic stem cell transplantation (HSCT) is the only curative treatment currently available and is carried out mostly in high risk patients. However, patients with low risk disease which fail first line treatments and present with life-threatening infections, severe anaemia/ thrombocytopenia and have poor risk molecular defects are also considered on a case by case bases (Platzbecker, 2019). Those who do not qualify for allogeneic HSCT continue with HMA therapy in particular Azacytidine treatment until disease worsens and allogenic HSCT is available. In all risk groups supportive care is critical for patients to alleviate symptom resulting from cytopenias, these include transfusions for problematic anaemia and thrombocytopenia along with antimicrobial treatments for febrile neutropenia's as infections are the leading cause of morbidity in MDS patients (Steensma, 2016). Excessive accumulation of iron is a common complication from continued transfusions in patients with MDS and so as part of the supportive care patients are treated with iron-chelating agents such as Deferasirox (Platzbecker, 2019).

Table 1.3: Therapies for the treatment of MDS

Symptom	**Agent**	
Thrombocytopenia	Thrombopoiesis-stimulating agents (TSA)	Eltrombopag
		Romiplostim
Anaemia	Erythropoiesis stimulating agents (ESA)	Epoetin
		Darbepoetin
Anaemia in genetically defined del(5q) MDS	Suppression of 5q clonal expansion	Lenalidomide*
	DNA hypomethylating agents (HMA)	Azactidine*
		Decitabine*
Immune dysregulation	Anti-T cell immunosuppressive therapy (IST)	Antithrombocyte globulin (ATG)
Excessive iron from transfusions	Iron-chelation agent	Deferasirox

*FDA-approved drugs:

1.2.2 Acute myeloid leukaemia (AML)

Acute myeloid leukaemia (AML) is the most common acute leukaemia in adults with a median age of 68 years at diagnosis and an incidence of ~1.3 cases per 100,000 in people younger than 65 years old to 12.2 cases per 100,000 in patients over 65 years old (I De Kouchkovsky and Abdul-Hay, 2016). It is an aggressive myeloid malignancy with vast heterogeneity characterised from failing haematopoiesis due to the accumulation of immature myeloid cells ('blasts') in the BM and PB (I De Kouchkovsky and Abdul-Hay, 2016), which involves abnormal proliferation and differentiation of a clonal population of myeloid stem cells. Most cases of AML occur *'de novo'* in previously healthy patients and causes are largely still undetermined. However, risk factors include age, smoking and exposure to chemical such as benzene and formaldehyde which account for a small proportion of patients. Patients with chronic MPN (Polycythemia vera, essential thrombocythemia and idiopathic myelofibrosis), MDS, genetic syndromes (Fanconi anaemia, GATA2 deficiency) and a family history of AML are also at risk of developing AML along with those who have undergone chemotherapy and high dose radiation exposure (Song *et al.*, 2018). The majority of patients present with a combination of leukocytosis and signs of BM failure such as anaemia and thrombocytopenia, reflecting the accumulation of poorly differentiated myeloid cells (termed 'blasts') within the BM, PB and less commonly in extramedullary tissues (I. De Kouchkovsky and Abdul-Hay, 2016). If left untreated patients die within weeks or months due to rapid expansion of blast cells which eventually replace normal haematopoietic cells and result in ineffective haematopoiesis. These blasts are present at >20% of the cell population in the BM, with the exception of CBF AML, NPM1 mutated AML or acute promyelocytic leukaemia which is diagnosed independent of blast count.

1.2.2.1 Molecular pathogenesis

The advent of next generation sequencing technologies has enabled the molecular profiling of patients with AML improving our knowledge of the heterogenous nature of this disease **(Table 1.4)**. Insights into mutational, chromosomal and clonal complexity have been highlighted along with the prominent amount of epigenetic defects that contribute to AML pathogenesis (Cai and Levine, 2019). A study by The Cancer Genome Atlas (TCGA) revealed that from the deconvolution of 200 AML genomes, a large proportion of patients had mutations in epigenetic regulators; DNA methylation-related genes (44%) and encoding chromatin modifiers or cohesion-complex genes (43%) (Ley *et al.*, 2013).

Table 1.4: Molecular events in AML leukemogenesis

Molecular alterations		Prevalence
Activation of pro-proliferation pathways	FLT3 mutations (ITD and TKD domain mutations)	28%
	N/RAS mutations	12%
	TP53 mutations	8%
	c-KIT mutations	4%
	Enhanced STAT3 signalling (cytokine induced or mutations in receptor tyrosine kinases; FLT duplications/ JAK2)	50% ⇩
Block of differentiation pathways	NPM1 mutations	27% ⇧
	CEBPA mutations	6% ⇧
Effects to both proliferation and differentiation	DNA methylation related gene mutations (DNMT3A, TET2, IDH-1 and IDH-2)	>40%

(⇧-favourable prognosis, ⇩poor prognosis)

At the molecular level AML development is largely associated with patients carrying chromosomal translocations such as AML1/ETO, CBFβ/SMMHC, MLL-related fusion genes and PML/RARα which result in chimeric protein formation altering the normal maturation process of myeloid precursor. These translocations are known as major initiating events of cell transformation which persistently involve TF that regulate normal haematopoiesis (Kellaway *et al.*, 2020). Chromosomal translocations involving the MLL gene and various different partners account for approximately 5-10% of human acute leukaemia's of which ENL, AF4 and AF9 are most predominant (Schwaller, 2020). These MLL- fusions are powerful oncogenes capable of transforming both HSCs and more committed progenitor cells, highlighted in studies employing retroviral overexpression of MLL-AF9 in murine GMP or HSPCs resulting in highly penetrant and transplantable leukaemia (Krivtsov *et al.,* 2006; Sommervaille and Cleary 2006). Nevertheless, additional co-operating mutations that drive leukemic growth such as activating mutations in common signalling pathway receptors are also required alongside other translocations. This is observed in RUNX1/ETO where although this translocation plays an important driver role in leukemogenesis through initiation of aberrant growth and differentiation phenotypes highlighted in conditional murine models, transgene expression alone is not sufficient to cause complete AML transformation (Rhoades *et al.*, 2000; Cabezas-Wallscheid *et al.*, 2013). This is further supported by the fact that the majority of t(8;21) AML patients present with additional mutations such as activating mutations in receptor tyrosine kinases (such as FLT3, RAS or KIT) that drive leukemic growth and are present in at least 30% of patients with AML (Kellaway *et al.*, 2020). Investigations into cooperating mutations with retrovirally transduced murine BM with RUNX1/ETO has highlighted the importance of these event in causing the development of AML in mice

(Grisolano *et al.*, 2003; Schessl *et al.*, 2005; Nick *et al.*, 2012). The heterogeneity of the AML is implicated through the interactions of chromosomal defects with the additional co-mutations. This has been observed in AML where c-Kit mutations are associated with t(8;21) or inv(16) which implicates good prognosis, along with Nuclear matrix protein 1 (*NMP1*) mutations observed frequently in patients with FMS-like tyrosine kinase 3-internal tandem duplication (*FLT3-ITD*) or in epigenetic genes DNMT3 and IDH1/2 (I De Kouchkovsky and Abdul-Hay, 2016). Whilst the majority of cases of AML are *de novo* around 4% of adults with AML carry inherited mutations in cancer susceptibility genes including: *CEBPA*, *GATA2*, *RUNX1* and others; Breast cancer gene 1 (*BRCA1*) and MutS Homolog 6 (*MSH6*), *DDX41* (Churpek, 2017)**.**

1.2.2.1 Diagnosis and treatment

Diagnosis is carried out by the assessment of clinical presentation and pathology from morphological evaluation of BM biopsy, PB smear, IHC, flow cytometry analysis and cytogenetic tests and molecular mutation analysis. This initial work up aims to firstly rule out differential diagnosis and identify AML and secondly to classify the subtype **(Table 1.9)** which aids in assessing risk-stratification which is vital when formulate treatment approaches such as whether intensive or reduced intensity chemotherapy regimens are required, along with decisions between consolidation chemotherapy or allogenic HSCT and choosing between established or investigational therapies (Narayanan and Weinberg, 2020). Reduced intensity (RI) treatments are important in elderly, those with comorbid conditions, advanced or refractory disease or those who have relapsed from autologous HSCT as these patients are unable to tolerate toxicity of intensive treatments. These RI are designed to reduce treatment related toxicity through reduced-dose cytotoxic conditioning whilst preserving anti-leukaemia affects. These patients are treated with a purine analogue such as fludarabine in combination with a second agent such as melphalan, idarubicin, busulfan, cytarabine or cyclophosphamide (Lazarus and Rowe, 2006). Cytogenetic changes have been observed to offer the strongest prognostic factors for complete remission (CR) and overall survival (OS) and cases have been further stratified based on cytogenetic profile alone **(Table 1.5)**.

Table 1.5: Prognosis- risk categories of AML based on cytogenetic profile (I. De Kouchkovsky and Abdul-Hay, 2016)

Prognostic risk group	Cytogenetic profile
Favourable	t(8:21)(q22;q22)
	inv(16)(p13;1q22)
	t(15;17)(q22;q12)
Intermediate	CN-AML
	t(9;11)(p22;q23)
	Cytogenetic abnormalities not included in the favourable or adverse prognostic risk groups
Adverse	inv(3)(q21q26.2)
	t(6;9)(p23;q34)
	11q abnormalities other than t(9;11)
	-5 or del(5q)
	-7
	Complex karyotype

Though significant advances have been made in the treatment of AML in younger patients, prognosis of the elderly, which accounts for the majority of new cases, remains poor (Shah *et al.*, 2013). These patients are more likely to present with adverse cytogenetic- risk profiles, are less likely to respond well to chemotherapy being more susceptible to treatment related toxicity. The first line of treatment is induction therapy which aims to achieve CR, though minimal residual disease (MRD) is commonly observed which results in relapse if treatment is discontinued. Standard induction therapy in cases with good to intermediate prognosis follows the '7+3' regimen; seven days of continuous infusion of cytarabine and three days of anthracycline. Patients with poorer prognosis such as those carrying DNMT3A and KMT2A mutations have been observed to benefit more from high dose daunorubicin treatment and FLAG-ID; fludarabine, cytarabine, G-CSF and idarubicin are used in cases of relapse. Therefore, an optimum treatment regime has yet to be define for these patients. Hypomethylating agents, which are usually used in the treatment in MDS have been trailed in AML patients and have reported effective results. In general, increased age and poorer-risk stratification groups are associated with lower rates of CR and decreased OS. The remission rates with 3+7 regime are 60-80% in younger patients (<60yrs) and 40-60% in older patients (>65yrs). Consolidation therapy is used to eradicate any residual disease, preventing relapse and aims to achieve long lasting remission, these treatments include chemotherapy and allogenic HSCT. Chemotherapy is offered to patients with favourable prognosis usually administered as intermediate dose cytarabine in 2-4 cycles of six doses each. Allogenic HSCT is offered as a first line treatment in patients with intermediate and most at risk AML which prolongs relapse free survival (RFS) and OS. There has been much investigation into novel

target therapies that can offer effective anti-leukemic activity with reduced toxicity from off target effects (I De Kouchkovsky and Abdul-Hay, 2016). The last several decades have seen very little changes to the intensive anthracycline-based induction regimes, followed with risk stratified post-remission therapy with cytarabine treatment or allogeneic stem cell transplantations. However, in 2017 four Food and Drug Administration (FDA) approved agents; Midostaurin, Enasidenib, Gemtuzumab ozogamicin and CPX 351 **(Table 1.6)** led treatment of AML into the age of precision medicine.

Table 1.6: Therapeutic agents for treatment of AML

Agent	Target	Description
Midostaurin (Rydapt)*	**Tyrosine kinase (TK) receptor inhibitors** FLT3 mutated 30% AML	Front line treatment of adults newly diagnosed with FLT3+ AML (ITD/TKD) In combination with cytarabine and daunorubicin induction and cytarabine consolidation(Stone *et al.*, 2017; Short *et al.*, 2019)
Gliteritinib*		Treatment relapsed or refractory FLT3-mutated AML (Short *et al.*, 2019)
Quizartinib		Treatment in relapsed/refractory AML (Currently in phase II clinical trials) (Short *et al.*, 2019)
C188-9	**STAT inhibitors** STAT3 tyrosine phosphorylation upregulation in >50% AML cases	(Redell *et al.*, 2011)
MM-206		(Krueger *et al.*, 2015)
OPB-31121		Clinical trials (Brambilla *et al.*, 2015)
Enasidenib (Idhifa)*	**IDH1/IDH2 small molecule inhibitors** Gain of function mutations in IDH-1/2 enzymes in 20% of AML cases	Treatment of adult patients with relapse or refractory AML with IDH2 mutations
Ivosidenib (Tibsovo)*		Treatment of adult patients with relapse or refractory AL (Golub *et al.*, 2019)
Gemtuzumab ozogamicin (GO) *	**Monoclonal antibodies** Direct ab-dependent cytotoxicity	Treatment of adults newly diagnosed CD33-positive AML and relapse or refractory CD33-positive AML in adults and paediatric patients. (Appelbaum and Bernstein, 2017)
*Glasdegib (Daurismo)	**Hedgehog signalling inhibitor**	Treatment of older (>75yrs) newly diagnosed AML: In combination with low-dose cytarabine (Hoy, 2019)
Pracinostat	**Histone deacetylase inhibitor**	Treatment of AML: In combination with azacytidine (Phase II clinical trial) (Garcia-Manero *et al.*, 2019)
Venetoclax*	**BCL2 inhibitor** Pro-apoptotic agents	Treatment of older patients newly diagnosed AML: In combination with low-dose cytarabine In combination with hypomethylation agents (decitabine or azacitidine) (Wei *et al.*, 2016; Pollyea *et al.*, 2019)
Clofarabine	**Nucleoside analogue** Interfere with DNA replication	Relapse or refractory paediatric ALL. Synergism with daunorubicin. Older patients with AML, in combination with decitabine (DNA methyltransferase inhibitor) (Tiley and Claxton, 2013; Van Eijkelenburg *et al.*, 2018)
CPX 351 (Vyxeos)*		Newly diagnosed Adults with t-AML or AML-MRC (60-70yrs) Liposomal daunorubicin/cytarabine (Alfayez *et al.*, 2020)

*FDA approved agents

1.2.3 Classification of myeloid malignancies

Classification systems for haematological malignancies are a criteria used to aid diagnosis into clinically relevant disease entities, which further support diagnosticians when choosing treatment regimes. The initial French-American-British (FAB) classification system was established in 1972 providing guidelines for diagnosis of MDS based on the lineages that are effected by the dysplasia and the percentage of blasts in the PB and BM and defined eight subtypes of AML based on morphological characteristics relying on histology examination of blast cells **(Table 1.7a and 1.7b*)***. With the advent of next generation sequencing technologies and its help advancing the fields understanding the molecular pathogenesis of both MDS and AML the world health organisation (WHO) introduced a new classification in 2001. This system was based on the FAB structure, but incorporated parameters describe the biology of the diseases including cell counts, dysplastic changes and genetic defects which were revised in 2008 and 2016 **(Table 1.8 and 1.9)**. The 2016 update of the WHO classifications featured for the first time 'Myeloid neoplasms with a germline predisposition' as a new provisional diagnostic entity, to apply cases of AML/MDS as a result of hereditary mutations causing susceptibility to myeloid malignancies. (Tefferi and Vardiman, 2008; Vardiman, 2012; Arber *et al.*, 2016; Czuchlewski and Peterson, 2016).

Table 1.7a: French-American-British (FAB) classification of MDS (Vardiman, 2012)

Category	Dysplasia	% BM blasts	% PB blasts
Refractory anaemia (RA)	Erythroid	<5	<1
Refractory anaemia with ring sideroblasts (RARS)	Erythroid	<5	<1
Refractory anaemia with excess blasts (RAEB)	2 or more lineages	5–20	0–4
Refractory anaemia with excess blasts in transformation (RAEB-T)	Usually 2 or more lineages	21–30	≥5
Chronic myelomonocytic leukaemia (CMML)	Variable ≥1 × 109/L monocytes	<20	

Table 1.7b: French-American-British (FAB) classification of AML

Class	Description
M0	Undifferentiated acute myeloblastic leukaemia
M1	Acute myeloblastic leukaemia with minimal maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute pro-myelocytic leukaemia
M4	Acute myelomonocytic leukaemia
M4EOS	Acute myelomonocytic leukaemia with eosinophilia
M5	Acute monocytic leukaemia
M6	Acute erythroid leukaemia

M7	Acute megakaryocytic leukaemia

Table 1.8: MDS Myeloid neoplasms WHO classifications (Arber *et al.*, 2016)

Classification	Subtype
Myeloproliferative neoplasms (MPN)	Chronic myeloid leukaemia (CML), *BCR-ABL1*$^{+}$
	Chronic neutrophilic leukaemia (CNL)
	Polycythemia vera (PV)
	Primary myelofibrosis (PMF) - PMF, prefibrotic/early stage - PMF, overt fibrotic stage
	Essential thrombocythemia (ET)
	Chronic eosinophilic leukaemia, not otherwise specified (NOS)
	MPN, unclassifiable
	Mastocytosis
Myeloid/lymphoid neoplasms with eosinophilia and rearrangement of *PDGFRA*, *PDGFRB*, or *FGFR1,* or with *PCM1-JAK2*	Myeloid/lymphoid neoplasms with *PDGFRA* rearrangement Myeloid/lymphoid neoplasms with *PDGFRB* rearrangement
	Myeloid/lymphoid neoplasms with *FGFR1* rearrangement
	Provisional entity: Myeloid/lymphoid neoplasms with PCM1-JAK2
Myelodysplastic/myeloproliferative neoplasms (MDS/MPN)	Chronic myelomonocytic leukaemia (CMML**)**
	Atypical chronic myeloid leukaemia (aCML), BCR-ABL1−
	Juvenile myelomonocytic leukaemia (JMML)
	MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T)
	MDS/MPN, unclassifiable
Myelodysplastic syndromes (MDS)	MDS with single lineage dysplasia
	MDS with ring sideroblasts (MDS-RS) - MDS-RS and single lineage dysplasia - MDS-RS and multilineage dysplasia
	MDS with multilineage dysplasia
	MDS with multilineage dysplasia
	MDS with isolated del(5q)
	MDS, unclassifiable
	Provisional entity: Refractory cytopenia of childhood
	Myeloid neoplasms with germ line predisposition

Table 1.9: Acute myeloid leukaemia (AML) and related neoplasms (Arber *et al.*, 2016)

Class	Subtypes
AML with recurrent genetic abnormalities	AML with t(8;21) (q22;q22.1); RUNX1-RUNX1T1
	AML with inv(16) (p13.1q22); CBFB-MYH11
	APL with PMLRARA- hypergranular
	APL with PMLRARA- microgranular
	AML with t(9;11) (p21.3;q23.3) children
	AML with t(9;11) (p21.3;q23.3)- adults
	AML with t(6;9) (p23;q34.1)
	AML with inv(3) (q21.3q26.2); GATA2, MECOM
	AML (Megakaryoblastic) with t(1;22) (p13.3;q13.1); RBM15-MKL1
	AML with BCR-ABL1
AML with gene mutations	AML with mutated NPM1
	AML with biallelic mutations of CEBPA
	AML with mutated RUNX1
AML with myelodysplasia-related changes (AML-MRC)	
Therapy-related AML (t-AML)	
AML, not otherwise specified (AML, NOS)	AML with minimal differentiation
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukaemia
	Acute monoblastic and monocytic leukaemia
	Acute erythroid leukaemia
	Acute megakaryoblastic leukaemia
	Acute basophilic leukaemia
	Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis
	Myeloid leukaemia associated with Down syndrome

1.3 Myeloid neoplasms with a germline predisposition

Though most cases of myeloid neoplasms are sporadic a small subset have been identified with germline mutations and therefore in 2016 the WHO revised the classification to include myeloid neoplasm (MDS, MDS/myeloproliferative neoplasms and AML) with a predisposing germline background. Though these familial cases are rare, the prevalence is likely to be underrepresented due to the heterogenous disease phenotypes and latency which is further impeded by the lack of family history (Tawana *et al.*, 2015). In these familial neoplasms the mutations are heterozygous and in the majority of cases result in loss of function alleles however, there gain of function mutants have also been described (Godley, 2014). It is widely appreciated that acquisition of secondary mutation or aberrant cytogenetics are necessary for progression to haematological malignancy. As genes which are mutated in familial predisposition towards myeloid malignancies are also found to be mutated in sporadic cases, investigation of familial myeloid neoplasms may also provide insight into the pathogenic mechanisms underlying disease progression.

Table 1.10: Predisposing haematopoietic transcription factors in familial myeloid neoplasms

Gene	Mutation frequency
CEBPA	**AML=** 1%
GATA2	**Childhood MDS**= 7% **Chronic neutropenia=** 6/14 pedigrees with other GATA2 features **Adult MDS/AML/CMML=** unknown
RUNX1	**MDS/AML=** unknown
ETV6	**Childhood ALL=** 10.8% **Inherited thrombocytopenia=** 2.6%
IKZF1	**Childhood B cell ALL=** 0.9%
PAX5	Only 2 Families reported

Modified from (Churpek and Bresnick, 2019a)

1.3.1 Familial platelet Disorder (FPD)- RUNX1

The first germline predisposition syndrome identified for myeloid malignancies was recognised in 1999, where germline heterozygous mutations in the RUNX1 gene caused a dominantly inherited familial platelet disorder (FPD), predisposing patients to acute myeloid leukaemia (Song *et al.*, 1999). FPD as a result of RUNX1 germline mutations is characterised by mild to moderate thrombocytopenia, along with aberrant platelet function and structure leading to poor wound healing and clinical bleeding during minor trauma or surgical procedures (Luo *et al.*, 2019). A spectrum of RUNX1 mutations have been described; frameshift and nonsense mutations have been reported throughout the gene, whilst missense mutations are focused in the highly conserved RUNT homology domain at residues important for DNA binding or heterodimerization. Partial and whole RUNX1 gene deletions have also been described along

with whole gene duplications, though these events are rarer events in patients (Churpek and Bresnick, 2019a). The molecular characteristics of different mutations have varying effects on the translated protein. Nonsense and deletion mutations have been reported to decrease RUNX1 protein levels, whilst missense mutations produce truncated proteins which are able to retain PEBP2-β/CBF-β dimerization, though lead to reduced or abolished DNA binding at the RUNX1-DNA interface. These mutations have also been observed to have dominant negative functions such as inhibition of WT RUNX1 expression (Michaud *et al.*, 2002; Matheny *et al.*, 2007). Whilst haploinsufficiency of RUNX1 causes FPD/AML in some cases with deletion and frameshift mutations, Michaud *et al* (2002) has suggested there to be an association between carrying a dominant-negative mutation and a higher prevalence of leukemic development, though further cohorts need to be investigated to solidify this idea. Furthermore, families carrying the same gene variant display heterogenous phenotypes with some members presenting with FPD/AML, whilst others remain asymptomatic, along with members of the same pedigree developing different subtypes of MDS/AML. The existence of asymptomatic carriers which display normal platelet counts throughout their lifetime suggests that FPD/AML has an incomplete penetrance. This highlights that one functional allele is adequate to support thrombopoiesis and that additional aberrant genetic events contribute to the disease phenotype (Owen *et al.*, 2008; Latger-Cannard *et al.*, 2016).

RUNX1 is known to act as a direct transcriptional activator of several proteins important for platelet function such as *PRKCQ, MYL9* and *ALOX12* and as a repressor of others MYH10 and ANKED26 (Jalagadugula *et al.*, 2010; Kaur *et al.*, 2010; Antony-Debré *et al.*, 2012; Bluteau *et al.*, 2014). Expression profiling of platelets from FDP/AML patients show downregulation of RUNX target genes such as ALOX12, MYL9, PLDN and MPL (Kaur *et al.*, 2010) involved in diverse platelet pathways. When these pathways are dysregulated, platelet abnormalities are observed which mirror those seen in human FPD/AML such as; decreased platelet aggregation and ATP secretion, dense and α-granule deficiencies and decreased platelet thrombopoietin receptors (Songdej and Rao, 2017). RUNX1 has also been shown to occupy target sites in ANKRD26, a gene encoding an ankyrin repeat protein which interacts with the inner cell membrane. Many mutations have been reported in t ANKRD26 and interestingly those located at RUNX1 binding sites at the 5’ UTR lead to another familial thrombocytopenia syndrome THC2 with a predisposition to myeloid malignancy very similar to FPD/AML. RUNX1 and ETS factor FLI1 function to supress ANKRD26 at these commonly mutated sites and failure to bind at these regions leads to increased thrombopoietin, MAPK and ERK signalling and defective proplatelet formation (Bluteau *et al.*, 2014). As ANKRD26 expression has been documented at elevated levels in FPD/AML patients, it has been

suggested that this network could also be implicated in the thrombocytopenia observed in FPD/AML.

The median age of onset to haematological malignancy in patients with FPD/AML is 33 years (5-79 years). Whilst MDS and AML occur most frequently, T-cell acute lymphoblastic leukaemia (ALL), non-Hodgkin lymphoma and other lymphoid malignancies have also been described (Owen *et al.*, 2008; Luo *et al.*, 2019). At MDS/AML presentation patients commonly have already lost the functional RUNX1 allele, have various chromosomal aberrations and additional somatic mutations have been described. These somatic mutations have been identified in CDC25C involved in cell cycle regulation, PDS5B involved in the cohesion complex during mitosis, epigenetic regulator TET2 and histone modifiers PHF6 and DNMT3A (Churpek and Bresnick, 2019a). While a great deal is known about the role of RUNX1 in megakaryopoiesis, it still remains unclear how haploinsufficiency leads to malignancy and how defects in platelet number and function are related to this progression. Interestingly, a study on a small cohort of FPD/AML patients revealed that 67% of asymptomatic patients younger than 50 years old display clonal haematopoiesis in the PB at a much higher proportion than the general population, suggesting that the baseline of mutagenic events is elevated in FPDMM HSPCs (Genovese *et al.*, 2014; Jaiswal *et al.*, 2014a; Churpek *et al.*, 2015). As there is a well-known link between RUNX proteins and the FA pathway involved in DNA repair, investigations into how genetic insults impact malignant transformation in FPD/AML might reveal novel mechanisms into disease progression (Wang *et al.*, 2014).

1.3.2 Thrombocytopenia 5 (THC5)- ETV6

Patients with thrombocytopenia 5 (THC5) have an autosomal dominant disorder due to heterozygous ETS Variant Transcription Factor 6 (ETV6) mutations. Patients present usually at childhood in a similar manner to patients with RUNX1 and Ankyrin Repeat Domain 26 (*ANKRD26*) mutations, with mild to moderate thrombocytopenia and a tendency to bleed (Feurstein and Godley, 2017). The syndrome has been observed to have complete penetrance, which suggest that one functional *ETV6* allele is not sufficient for normal platelet development. In support of this, functional studies have revealed that patient-derived megakaryocytes show defective proplatelet maturation and decreased polyploidization during *in vitro* culture however, platelet defects of aggregation, activation or adhesion were only observed in a small proportion of patients and flow cytometry analysis revealed no defects to glycoproteins on platelet surface (Melazzini *et al.*, 2016). Furthermore, patient bone marrow analysis commonly reveals small hypoblated megakaryocytes, mild dyserythropoiesis and mild myeloid dyspoiesis (Galera, Dulau-Florea and Calvo, 2019).

The *ETV6* (TEL) gene is located on chromosome 12p13.2, contains eight exons and is essential in the development of the embryo (Hock *et al.*, 2004). It is well known as an important regulator of haematopoiesis since its first account as a translocation partner of Platelet Derived Growth Factor Subunit B (PDGFB) in chronic myelomonocytic leukaemia (Golub *et al.*, 1994). Since this first report ETV6 has been described as one of the most commonly re-arranged genes in myeloid and lymphoid malignancies with more than fusion partners (Churpek and Bresnick, 2019a). In contrast to other activating ETS transcription factors, ETV6 acts as a transcriptional repressor that binds DNA via a C-terminal DNA binding domain, which is vital for autoinhibition and homodimerization which is highly conserved in ETS-family transcription factors. The N- terminal Pointed (PNT) domain also known as helix loop helix (HLH) or sterile alpha motif (SAM), mediates dimerization and acts as a corepressor docking site. A 'linker region' located between these two domains acts as a autoinhibitory domain, restricting ETV6 DNA binding (Hollenhorst, McIntosh and Graves, 2011; Churpek and Bresnick, 2019b).

Germline mutations involved in THC5 are mostly missense and located within the C-terminal domain, though one recurrent linker mutation (p.P214L) has been described along with rare truncating or missense mutations outside of these two domains. Many studies have revealed that ETV6 mutations have been reported to decrease DNA binding, cause aberrant nuclear localisation leading to cytoplasmic localisation and reduce transcriptional repression in a dominant negative fashion (M. Y. Zhang *et al.*, 2015; Melazzini *et al.*, 2016; Galera, Dulau-Florea and Calvo, 2019). A large proportion of THC5 patients go on to progress toward haematological malignancies of both myeloid and lymphoid lineages. Approximately 30% of patients with ETV6 mutations are diagnosed with haematological malignancies such as biphenotypic acute leukaemia (BAL), MDS, AML, CMML and Polycythaemia Vera with *JAK2* mutations. These patients present at a younger age than in patients with myeloid malignancies without ETV6 mutations and has been suggested to be associated with the acquisition of further somatic mutations. A further 20% develop childhood-onset B-ALL accounting for the highest frequency (Churpek and Bresnick, 2019b; Galera, Dulau-Florea and Calvo, 2019).

1.3.3 Familial AML- CEBPA

Acute myeloid leukaemia is predominantly a sporadic disease however, germline CEBPA mutations, though rare, are the major cause of familial AML. The prevalence of familial CEBPA AML has been estimated from a study of 187 patients with AML revealing 1% are germline CEBPA mutated and 11% are both somatic and germline (Pabst *et al.*, 2008). Familial AML is inherited in an autosomal dominant fashion and has complete or near complete penetrance

with a generally favourable prognosis in comparison to normal karyotype AML (survival rates around 50-60% compared to 25-40%) (Renneville *et al.*, 2008; Churpek and Bresnick, 2019b).

The majority of causative mutations in familial CEBPA are frameshifts found within the N-terminal which utilise the second transcriptional start site in the CEBPA mRNA, resulting in the expression of the shorter isoform p30. This isoform lacks two of three trans-activation elements along with the C-terminal basic region leucine-zipper (bZIP) required for DNA binding and dimerization. Both the longer p42 and shorter p30 isoforms homo or heterodimerise with CEBPA family members and regulate the expression of genes involved in cell grown, survival, differentiation, metabolism and inflammation and their ratio of expression determines cell fate as high levels of p42 blocks proliferation resulting in differentiation whilst high p30 expression is associated with an immature cell state and inhibition of terminal differentiation towards adipocytes and neutrophils (Calkhoven, Müller and Leutz, 2000; Schmidt, Heyes and Grebien, 2020). Therefore, mutations perturb the isotype ratio reducing p42 expression available for promoting cell differentiation and cell cycle arrest causing dominant-negative activity (Kirstetter *et al.*, 2008; Pulikkan, Tenen and Behre, 2017).

As with other myeloid malignancy predisposition genes, single germline CEBPA mutations alone are not sufficient to induce AML as many families reported with CEBPA mutations are clinically normal until emergence of AML (Tawana *et al.*, 2017) . At the time of AML development most patients are reported to have acquired a secondary CEBPA mutation in the wild-type allele, leading to the disruption of the C-terminal leucine zipper, suggesting an independent leukemic episode arising from the highly penetrant germline mutations. This has also been supported by findings in mouse models where animals with heterozygous frameshift mutations failed to develop leukaemia, though animals with both frame shift and missense mutants lead to an accelerated disease development (Bereshchenko *et al.*, 2009).

Familial and sporadic CEBPA mutations share similar AML pathological presentations with a predominance of FAB subtype M1, M2 or M4 morphology, abnormal eosinophils and normal cytogenetics. Abnormalities such as the presence of Auer rods in myoblast cells and aberrant CD7 expression on leukemic blasts have also been described in patients. Along with an acquired secondary CEBPA mutation, other additional somatic mutations that have been documented include GATA2, WT1 and EZH1 are the most common (Churpek and Bresnick, 2019b). This uniform AML phenotype differs from the other transcription factor mutation predisposition syndromes as it parallels the phenotype of CEBPA dysregulation in a more lineage restricted manner as opposed to the broader spectrum of haematopoietic-regulatory activities of RUNX1 and ETV6 (Churpek and Bresnick, 2019b).

1.3.4 GATA2 myeloid malignancy

To date more than 100 GATA2 mutant alleles, in approximately 400 cases have been described throughout the literature **(Figure 1.5).** One third of these mutations are inherited and are therefore the initiating mutation whilst the remaining occur *de novo* in association with other driver mutations as secondary events (Collin, Dickinson and Bigley, 2015a).

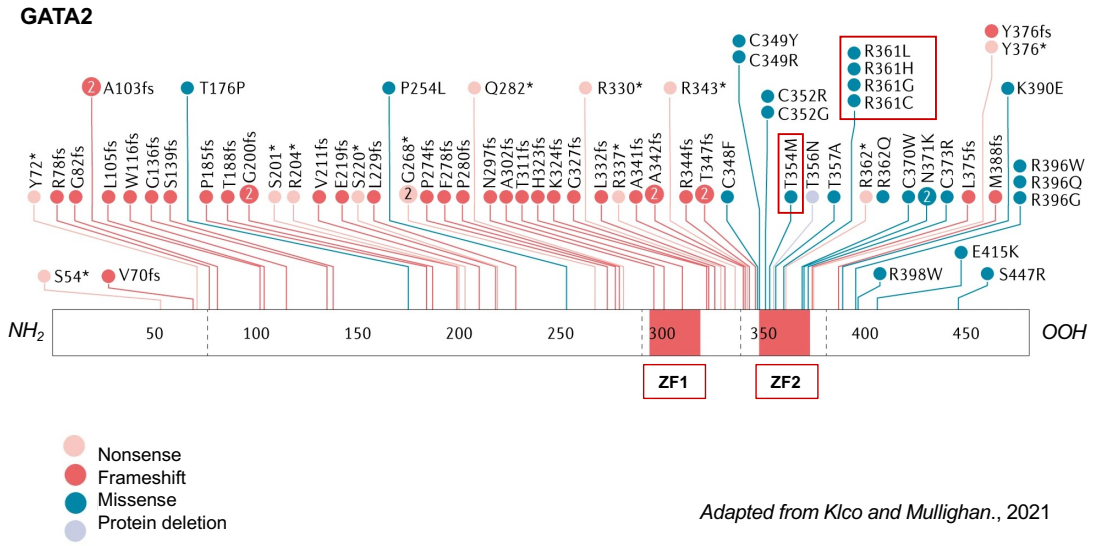


Figure 1.5. Germline and sporadic GATA2 mutations. Over 100 mutations have been documented throughout the GATA2 gene. A representation of some of the *GATA2* mutations found in hereditary *GATA2* syndromes and in myeloid malignancies and their location in the gene. The number indicates the amino acid altered and the colour representing the type of mutation. Red boxes highlight prevalent germline GATA2 mutations observed in patients modelled in this thesis.

Mutations in the CCAAT/enhancer binding protein a (CEBPA) gene are found in patients with cytogenetically normal (CN) AML. Patients carrying biallelic mutations typically have a combination of a N-terminal frameshift mutation, leading to a 30-kDa dominant negative CEPBA isoform and a C-terminal in-frame mutation in the bZIP region, which disrupts dimerization and DNA binding activity (Greif *et al.*, 2012). Only patients with biallelic mutations have a favourable outcome in comparison to other cytogenetic normal (CN)-AML patients (Konstandin *et al.*, 2018).

GATA2 mutations have been identified in high frequencies in these patients which are mostly of a granulocytic (M1 or M2) subtype, though mutations have also been observed in patients with acute erythroid Leukaemia (AEL) subtype M6 (Ping *et al.*, 2017). These GATA2 mutations are located in the ZF1 domain and are secondary events (Fasan *et al.*, 2013). The acquisition of the *GATA2* mutation in these patients is thought to be an early event in leukemogenesis,

supported by the highly correlated mutant levels of both GATA2 and CEBPA (Green *et al.*, 2013). It has been suggested that considering the direct protein-protein interaction of GATA2 and CEBPA and their high expression in early HSPCs, mutations in both genes may affect the same protein complex in a synergistic manner during haematopoietic differentiation leading to malignant transformation. This is supported by the observation that *GATA2* mutations in BiCEBPA patients had a reduced capacity to cooperate with wild-type CEBPA to activate transcription (Greif *et al.*, 2012). Though some studies have shown additional *GATA2* mutations in BiCEBPA mutant patients have a further favourable impact on patient survival (13/33) (Fasan *et al.*, 2013) others have reported no impact (15/55) (Green *et al.*, 2013) therefore, further investigation on larger cohorts maybe necessary to draw any final conclusions.

The *GATA2* somatic mutation L359V affecting the C terminal zinc finger (ZF2) is associated with the progression of chronic myeloid leukaemia (CML), a malignant clonal disorder of HSPCs accounting for 15% of all adult leukaemia's. Various studies have identified BCR/ABL as the major cause of CML (Ramaraj *et al.*, 2004). GATA2 L359V plays an important role in the progression of CML, identified in 10% of CML patients at progression phase (AP & BC) however, it has not been detected in patients at the chronic phase (CP) highlighting it as a secondary event (Zhang *et al.*, 2008). Large scale screening of patients with AML, MDS, ALL, CLL, CML (CP) and BCL/ABL negative MPD have shown the L359V GATA2 mutation to be an exclusive secondary event only in the progression of CML to blast crisis (BC) (Zhang *et al.*, 2009).

1.3.4.1 Germline GATA2 mutations

GATA2 germline mutations were first observed in 2011 in various families with MDS/ AML and immunodeficiency syndromes. It is now understood that these GATA2 mutations are initiating events which cause an immunodeficiency syndrome with a predisposition to MDS and AML. Consequently, GATA2 now joins a group of myeloid neoplasms with a germline predisposition along with *CEBPA*, *DDX41*, *RUNX1*, *ANKRD26* and *ETV6* according to the 2016 WHO classification.

Most reported germline mutations have been described within the C-finger and fall into one of four categories; truncating mutations (proximal to the C-finger), missense mutations (within the C-finger), noncoding mutations (within the '+9.5' enhancer, located within intron 4 and -77 enhancer mutation) and mutations resulting in aberrant mRNA splicing (5' (c.1018-1G>T), 3' exon-intron boundary of exon 5 (c.1143+5G>A) (Dickinson *et al.*, 2011)(Schlums *et al.*, 2017) and c.351C>G in exon 3 (Wehr *et al.*, 2018)). Nevertheless, recently a novel mutation has

been reported; an in-frame insertion mutation p.A345delinsALLVAALLAA at exon 5. This mutation results in the increased separation of the first and second zinc fingers by a 9 amino acid insertion therefore defining a new class of germline mutation (Cavalcante de Andrade Silva *et al.*, 2020).

Pathological GATA2 mutants commonly disrupt amino acid residues which mediate DNA binding or *cis* elements in vital GATA2 intronic enhancers. This has suggested a haploinsufficient mechanism of pathogenesis (Hahn *et al.*, 2011; Cavalcante de Andrade Silva *et al.*, 2020). Nevertheless, it is important to note that not all pathological GATA2 mutants are strictly inhibitory as demonstrated in complementation assays using the GATA2 -77 enhancer mutant, which showed an increase in myelopoiesis-promoting activity compared to wild-type levels (Katsumura *et al.*, 2018). This reveals that the haploinsufficiency paradigm does not fully explain GATA2 linked pathogenesis and there could be a combination of both quantitative and qualitative defects that occur due to these mutations, which could help explain the complex phenotype observed among patients (Katsumura *et al.*, 2018).

1.3.4.2 GATA2 immunodeficiency syndromes

Four presentations of GATA2 patients carrying germline heterozygous mutations have been described, each initially thought to be individual disease entities; MonoMac syndrome, dendritic cell, monocyte, and lymphocyte (DCML) deficiency, Emberger syndrome and familial MDS/AML (Table 11). These syndromes are now recognised as different manifestations of a single genetic disorder with a broad phenotype- GATA2 deficiency (Spinner *et al.*, 2014). The clinical hallmarks include immunodeficiency with increased susceptibly to human papilloma viruses (HPVs), nontuberculous mycobacteria (NTM), pulmonary alveolar proteinosis (PAP), congenital lymphedema and predisposition to MDS /AML (Spinner *et al.*, 2014). Vast clinical heterogeneity has been described between patients with GATA2 deficiency, with age at disease onset ranging from early childhood to late adulthood and clinical presentations varying from life-threatening infections and leukaemia to patients presenting as asymptomatic carriers (Spinner *et al.*, 2014) (see **Table 1.12** for all documented clinical presentation to date).

Table 1.11: GATA2 immunodeficiency syndromes

Syndrome	Features	Cytopenias
MonoMac	Mycobacterial infection	monocytes
	Human papilloma virus (HVP) infection	
	Fungal infection	
DCML deficiency	Mycobacterial infections	Dendritic cell
	HVP	monocytes
	Pulmonary alveolar proteinosis	B cells
		NK cells
Emberger syndrome	Primary Lymphoedema	T cell <CD4/CD8 ratio
Familial MDS/AML	With no previous malignancy	

1.3.4.2.1 Clinical manifestation/phenotype

Individuals carrying germline GATA2 mutations appear immunologically and haematologically normal, as observed in patients who have been diagnosed at birth through genetic screening due to an affected family member. Over their lifetime these patients develop cytopenias where there is significant variability. However, circulating monocytopenia, B and NK cell lymphocytopenia and dendritic cell depletion seem to be characteristic features of GATA2 immunodeficiency from peripheral blood analysis.

It has been highlighted from historical PB counts over several years in DCML patients that monocytopenia persists from 3-12 years in some cases, though seems to be a feature associated later in disease development (Bigley *et al.*, 2011; McReynolds, Calvo and Holland, 2018). With the exception of patients that present with aplastic anaemia (AA), MDS or AML, anaemia and thrombocytopenia are uncommon early presentations that are observed in various other Bone Marrow Failure (BMF) syndromes.

Patient bone marrow analysis is routinely carried out during diagnosis, though histological, cytological and cytogenetic examination of the bone marrow can reveal normal haematopoiesis in GATA2 deficient patients. Bigley *et al* (2011) and colleagues reasoned that as patients with DCML deficiency have normal erythroid, granulocyte and platelet production this disorder could be linked to a specific progenitor defect. Upon in-depth flow cytometry analysis of three patients with DCML deficiency the authors revealed; a complete absence of myeloid lymphoid progenitors (MLP) marked by $CD38^{-}CD90^{-/lo}CD45RA^{+}$, a significant reduction in granulocyte macrophage progenitors (GMP) marked by $CD38^{+}CD10^{-}CD45RA^{+}$ (72-95%) and the abolishment of B and NK precursors ($CD38^{+}CD10^{+}$) which are the

immediate progeny of the MLP. The same findings have been observed in patients presenting with Emberger syndrome, MonoMac and familial MDS associated GATA2 mutations (Dickinson *et al.*, 2014). This reveals that the immunodeficiency observed in GATA2 deficient patients arise from a defective bone marrow with a specific defect in the MLP and GMP compartments resulting in the hallmark cytopenias of monocyte, NK, B, DC cell lineages, whilst the MEP compartment is preserved allowing for normal erythrocyte and megakaryocyte production in patients. It is important to note that a sufficient amount of GMPs are preserved which are able to maintain granulopoiesis, though is insufficient to support monocyte and DC development (Bigley *et al.*, 2011; Dickinson *et al.*, 2014).

Table 1.12: Clinical heterogeneity reported among GATA2 deficient patients(Spinner *et al.*, 2014)**.**

Immunological	- B lymphocytopenia - NK lymphocytopenia - **Monocytopenia*** - CD4 lymphocytopenia - Neutropenia - Pancytopenia (MDS patients) - Monocytosis (CMML patients)
hematologic	- MDS 84% - AML 14% - chronic myelomonocytic leukaemia - LGL lymphoproliferation - leukaemia 8%

Infectious	- Viral infections:
	  ○ Human papilloma viruses (HPVs)
	  ○ Hepatitis C virus (HCV)
	  ○ Herpes simplex virus (HSV)
	  ○ Varicella zoster virus (VSV)
	  ○ Cytomegalovirus (CMV)
	  ○ Epstein–Barr virus (EBV)
	- Bacterial infections:
	- Mycobacterium avium complex
	  ○ *Mkansasii* } Slow growing
	  ○ *Mszulgaiand*
	  ○ *Mfortuitum* } Rapid growing
	  ○ *Mabscessus*
	  ○ *M chelonae*
	- fungal infections 16%

Pulmonary	- Diffusion defects 79% - Ventilatory defects 63%, - Pulmonary alveolar proteinosis (PAP)18% - Pulmonary arterial hypertension 9% - Pneumonia
Dermatologic	- Warts 53%, - Panniculitis 30%
Neoplastic	- Human papillomavirus1 tumours 35% - Epstein-Barr virus1 tumours 4%
Vascular/lymphatic	- Venous thrombosis 25% - Congenital lymphedema 11%
Other	- Sensory-neural hearing loss 76% - Miscarriage 33% - Premature Labour - Hypothyroidism 14%. - Otitis - Sinusitis
Asymptomatic carriers	

*monocytopenia has been shown to be the accessory phenotype of these patients later in disease development

It has been observed that monocytopenia leads to a great impairment of whole blood cytokine responses in screening tests for genetic susceptibility to mycobacterial infection. It has been outlined that profound defects in the production or signalling of type I cytokines such as IFN-γ, IL-12, 18, 23 and 17 lead to a predisposition to infections from weakly pathogenic mycobacteria. During the hosts interaction with an intracellular pathogen, macrophages undergo activation through recognition of the micro-organisms Pathogen-Associated Molecular patterns (PAMP) through their Toll-like receptors (TLRs), triggering cytokine production. Along with TNF-a and IL-1b, which play key roles in granuloma formation and reducing bacterial growth, IL-12 secretion during endocytosis of mycobacteria is vital in bridging the gap between innate and adaptive immunity by inducing activation and differentiation of T cells into T-helper 1 cells ($CD8^+$) which secrete IFN-y (Ramirez-Alejo and Santos-Argumedo, 2014). Therefore, the IL-12/IFN-y axis is very important in the control of infections by intracellular pathogens. This could explain the increased susceptibility to bacterial infections observed in patients with GATA2 deficiency, supported by the clear monocytopenia, along with the reported impairment of IL-12 and IFN-y responses in DCML patients by Bigley *et al* (2011).

As mentioned above B/NK precursors ($CD38^+CD10^+$) are abolished in the BM of GATA2 deficient patients along with the complete or near depletion of transitional B cells ($CD38^+CD27^-$). Naive B cells have also been reported low in abundance whilst surprisingly there is a relative enrichment of memory B cells and plasmablast observed in the BM and at sites of inflammation suggesting a skew towards a mature phenotype. Furthermore relatively normal immunoglobulin (Ig) levels are reported in most patients. (Dickinson *et al.*, 2014; Collin, Dickinson and Bigley, 2015a). It has been suggested that these mature cells represent local persistence and proliferation from previously produced cells (Vinh *et al.*, 2010).

Additionally, the most immature population of NK cells ($CD56^{bright}$) are absent along with a skew towards a more differentiated phenotype in the $CD56^{dim}$ population shown by the loss of NKG2A and CD62l and expression of killer-cell immunoglobulin-like receptors (KIR) (Béziat *et al.*, 2010; Dickinson *et al.*, 2014). This can be related to the requirement for GATA2 for normal and complete maturation of NK cells along with the $CD56^{bright}$ population along with full cytotoxicity(Mace *et al.*, 2013). A small proportion of patients reported with absent NK precursors have been documented with normal overall NK counts, how this occurs is yet to be defined, though the development of 'NK memory' as a result of chronic antigen exposure

has been suggested (Romee *et al.*, 2012; Dickinson *et al.*, 2014) . This again suggest the potential of long-lived NK cells which persist after the depletion of NK precursors though these features do vary between patients.

Conversely, T-cells are relatively well conserved in GATA2-deficient patients. Due to the fact that T-cell homeostasis in adults is independent of BM-derived T-cell precursors, patients with BMF syndromes T-cells along with plasma and Ig can be maintained for several years. The inversion of the CD4:CD8 ratio has been identified as a rough sign that CD4 helper function is failing in patients and that expansion of CD8 memory cells is being driven by chronic antigen stimulation (Vinh *et al.*, 2010; Ostergaard *et al.*, 2011; West *et al.*, 2014). In support of this notion further immunophenotyping of the T-cell subsets in patients revealed a decrease in naïve cells and an expansion of terminally differentiated effector cells ($CD8^+CD45RA^+$) (Dickinson *et al.*, 2014). Taken together, immunophenotyping of the lymphoid compartment in GATA2 patients mirrors the terminal differentiation observed in patients with advancing age and chronic viral infections and therefore a prematurely aging phenotype.

The loss of DC in patients with GATA2 deficiency was first described during the initial discovery of DCML (Bigley *et al.*, 2011) and is thought to impair the recognition of viral and intracellular pathogens which contributes to the disseminated infections observed in patients. Patients with no detectible peripheral DCs are still at risk of graft-versus-host disease upon transplantation due to the sustained populations of mature B-cells that are able to activate T-cells (Cuellar-Rodriguez *et al.*, 2011) **(Figure 1.6)**.

Table 1.13: Prevalence of GATA2 deficiency related myeloid neoplasms (modified from Wlodarski, Collin and Horwitz, 2017)

Screening cohort	GATA2-mutated cases	Cases with myeloid neoplasms	Age at diagnosis myeloid neoplasm (Median)years	Monosomy 7 and der(1:7) (%)	ASXL1 -7	Trisomy +8 (%)	Somatic mutations	Reference
Familial MDS/AML	28	15	10-53 (20.5)	50		10		Hahn- 2011
Emberger syndrome	14	8	9-53 (12)	75		0		Ostergaard-2011
Familial MDS	5	2	18-23 (20.5)	2/2		0	**ASXL1 2/2**	Bodor-2012
MDS/AML	10	9	10-33 (16)	44		11		Kazenwadel-2012
Familial MDS	4	4	12-48 (20)	25		0		Holme-2012
Chronic neutropenia	14	10	6-35 (15)	70		0		Pasquet-2013
MonoMac	32	20	3-78 (21.5)	n.a		n.a		Hsu-2013
GATA2 deficiency	57	42	0.4-78 (19)	21		24		Spinner-2014
GATA2 deficiency	48	42	12-78 (35.5)	20	6	22.5	**ASXL1 14/42**	West-2014
GATA2 deficiency	30	11	4-40 (25)	18		18		Dickinson-2014
Myeloid neoplasm	5	3	7-38 (22.5)	0		0		Mir-2015
GATA2 deficiency	28	28	14-60 (30)	23		23		Ganapathi-2015
Familial MDS	7	6	13-68 (16.5)	0		25		Churpek-2015
Paediatric MDS	6	5	n.a	2/6 80	1/6	3/6 20	**ASXL1 2/6** RUNX1 1/6, STAG2 1/6 TP53 1/6 SETBP1 1/6 NRAS 2/6	Wang-2015
Paediatric MDS	5	5	12-22 (16)	0		40		Zhang-2015
Paediatric MDS	60	57	3-19 (12)	75		9		Wlodarski-2016
Paediatric MDS	12	10	4.4-17 (14.5)	80		10		Novakova-2016
GATA2 deficiency	13	5	7-60 (18)	n.a		n.a		Schlums-2017
Paediatric MDS	15	15		2/15	2/15	0	SETBP1 6/15 RUNX1 2/15 **ASXL1 2/15**	Pastor 2016
Familial MDS	4	2	17-18	0	0	1	STAG2 2/2	Ding 2017
Paediatric MDS	5	5	5-15 ()	4/7	0/7	0	RUNX1 1/5, SEPTBP1 1/5, IKZF1 1/5, CRLF2	Fisher, 2017
GATA2 deficiency	79	57	7-72 (	27/79	n.a	16/79 (12.6)	n.a	Donadieu, 2018
GATA2 deficiency	7	4	9-56 (	1/7	1/7	1	**ASXL1 3/4** DNMT3A 1/4, ETV6 1/4	Haddox, 2018

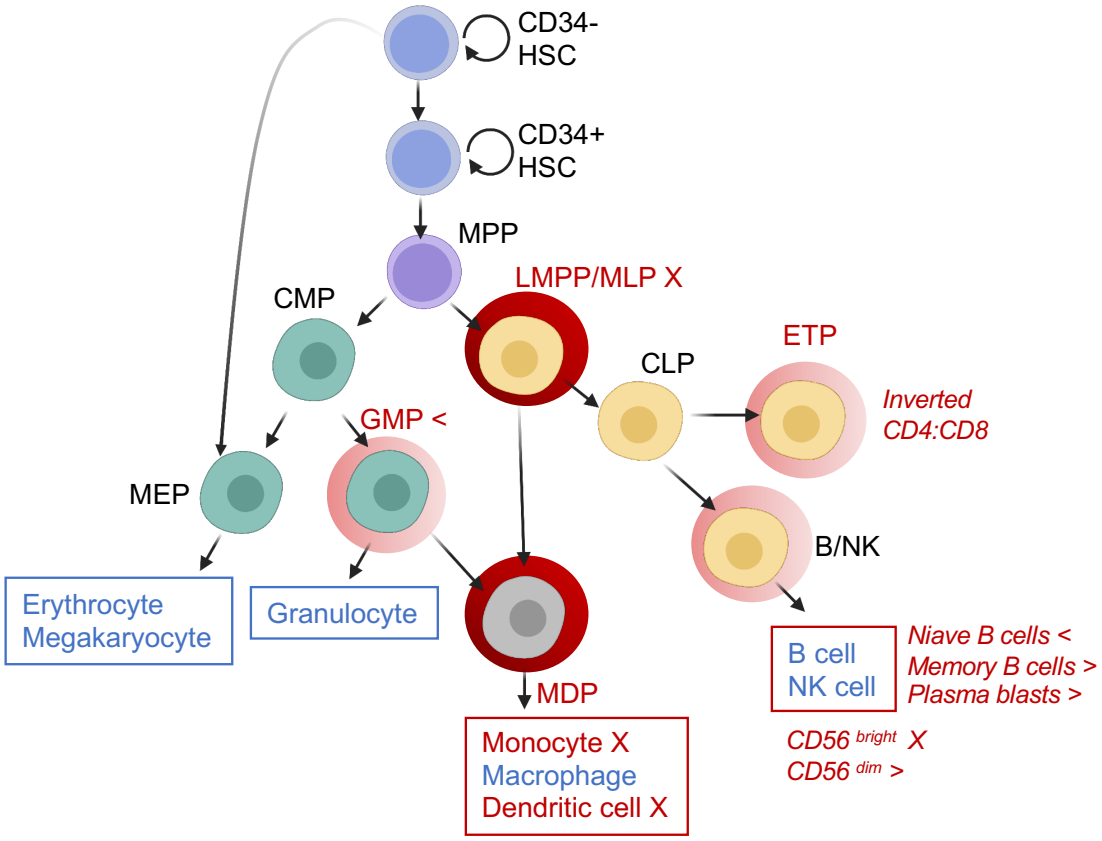


Figure 1.6: Lineages documented to be affected in patients harbouring GATA2 germline mutations. Vast clinical heterogeneity has been reported among GATA2 deficient patients. Bone marrow analysis has revealed a complete absence of myeloid lymphoid progenitors (MLP) marked by CD38-CD90-/loCD45RA+, a significant reduction in granulocyte macrophage progenitors (GMP) marked by CD38+CD10-CD45RA+ (72-95%) and the abolishment of B and NK precursors. Naive B cells are reduced with an enrichment of memory B cells and plasmablast and the most immature population of NK cells ($CD56^{bright}$) are absent with differentiation skewed towards mature cells ($CD56^{dim}$). Though T cells are preserved there in an inversion of CD4:CD8 ratio. Dendritic cells (DC) and monocytes are often depleted.

1.3.4.2.2 Clinical management

Currently, the only curative treatment for patients with GATA2 deficiency is HSCT which is highly effective at restoring normal haematopoiesis, resolving MDS, clearing underlying infections and resolving pulmonary hypertension and PAP. Ideal timing of transplantation is unclear, in general if cytopenias are present or if life-threatening complications such as infections or organ dysfunction develops patients proceed to transplantation. However, disease progression is difficult to predict, and many patients are diagnosed during later

disease stages such as MDS/AML. GATA2 deficient patient screening and monitoring varies though typical parameters are identified in **Table 1.14**, if MDS or AA develops patients are referred for immediate HSCT. It is important to note that sequencing of the GATA2 locus remains challenging as mutations have been identified throughout the whole region and complete coverage of the gene for mutation screening requires extensive optimisation. Moreover, when patients present with GATA2 deficiency there are often cases where sequencing of exons and intron 5 are negative, in these circumstances copy number variation analysis and mRNA sequencing is necessary to screen for larger deletions and splicing mutations (Bogaert *et al.*, 2020). Family member potential donors are screened for GATA2 mutations along with HLA typing, which is vital giving the vast clinical heterogeneity of the GATA2 phenotype and age at presentation (McReynolds, Calvo and Holland, 2018).

Table 1.14: GATA2 patient monitoring and screening (Evaluated yearly unless further stated).

Screening	GATA2 mutations identified
	Cytopenias identified
	Baseline BM identified Biopsy and aspiration 1. Mycobacterial culture 2. Flow cytometry analysis 3. Cytogenetics 4. Somatic mutation panel testing
Monitoring	Vaccinations updated
	Daily antibiotic treatment (azithromycin) prophylaxis (NTM)
	Complete blood count (CBC) 1. Differential- Biannually 2. Lymphocyte subsets- Annually
	Pulmonary function test
	Skin examination
	Gynaecological examinations (Female)

In clinical practise, guidelines on conditioning regimes, donor and stem cell source for HSCT along with GVHD prophylaxis is currently lacking. Many groups have reported good outcomes of HSCT in in GATA2 deficient patients with both myeloablative and reduced intensity preparative regimens. Moreover, reports of successful matched related and unrelated PBMCs, matched and haploidentical BMs and UC grafts in a variety of different disease outcomes such as disseminated HPV skin disease and Emberger syndrome with MDS and AML have been described (Maeurer *et al.*, 2014; Lübking *et al.*, 2015; Ciullini Mannurita *et al.*, 2016; Mallhi *et al.*, 2016; Ramzan *et al.*, 2017). GVHD has been observed at normal and

increased rates in GATA2 deficient patients however, differences in prophylaxis treatments between cases make it difficult to conclude GVHD rates (McReynolds, Calvo and Holland, 2018). Interestingly a report of monozygotic twins underwent HSCT from the same donor, with the same conditioning regime however, one developed sever acute GVHD (skin and gastrointestinal tract) requiring long term immunosuppression, whilst the second developed grade 1 GVHD (skin) requiring systemic corticosteroids (Kim *et al.*, 2019). This case study highlights nonheritable factors that affect GVHD outcomes such as type of infection and timing in relation to transplant procedure (prior/during).

1.3.4.2.3 Incomplete penetrance

Variable or incomplete penetrance is a phenomenon whereby not every individual carrying the same variant in a disease-causing gene presents in the same manner and is a commonly observed among families carrying germline GATA2 mutations. Work conducted by Seraihi et al (2018) investigated whether the varied disease symptoms observed in GATA2 families were modulated by endogenous levels of GATA2 expression. Interestingly, the authors reported GATA2 levels from a T354M GATA2 pedigree were significantly lower in symptomatic carriers (IV-10) compared with an asymptomatic carrier (III.7). Furthermore, the relative expression of endogenous GATA2 was reduced upon initial analysis when symptoms of monocytopenia and neutropenia had begun (IV-10 year 1) however, after the patient's condition had stabilised, GATA2 levels increased by two and three-fold respectively in the following years (year 4 and 6) suggesting re-activation of the WT allele (Al Seraihi *et al.*, 2018). Sanger sequencing revealed allele specific expression (ASE) of the mutant allele, with absence of the wildtype in the symptomatic patient, in contrast with bialleic expression of both mutant and wild-type expression in asymptomatic carriers. These findings suggest an allelic imbalance in the ratio of wildtype:mutant GATA2 expression, which may explain the variable disease penetrance observed in pedigrees carrying the same GATA2 mutation documented extensively in the literature. This is thought to be one explanation of why inherited diseases are transmitted through un-effected parents, how clinically healthy or 'asymptomatic' family members are able to carry pathogenic variants of a gene, yet do not present with features of disease and why family members of carrying the same mutation variant can display variable clinical manifestations (Taeubner *et al.*, 2018).

Reduced penetrance presents a big challenge in the diagnosis and treatment of these families, especially during genetic counselling where potential stem cell transplant donors are screened. Several reports have observed donor derived MDS/AML in families with germline GATA2 mutations. Galera *et al* (2018) reported haploidentical allogeneic HSCT transplantation of a T354M GATA2 positive patient (21years old) presenting with AML with

MDS-related changes. The mother of the patient (64 years of age) was chosen as a donor candidate as there was no previous history of haematological or immunodeficient disorders. Sadly, two years post HSCT the patient developed neutropenia and cytogenetic studies revealed a donor karyotype of monosomy 7 and trisomy 8. Upon further analysis, sequencing identified a GATA2 mutation (T354M) in both parent donor and recipient samples. Three years post HSCT the patient was diagnosed with donor derived MDS (Galera *et al.*, 2018). Identification and exclusion of related asymptomatic carriers is vital in GATA2 patients as their use in HSCT can have detrimental effects on patient outcome.

The variations of GATA2 gene expression highlighted in the above studies suggest a role of differential epigenetic regulation between patients. Promoter methylation, small interfering RNAs and histone acetylation are all responsible for regulating gene expression and could therefore have an impact on disease development (Kim *et al.*, 2019). In support of this a case study following monozygotic twins harbouring the same GATA2 mutation (R398W) though showing completely different disease phenotypes was investigated for differences in methylation patterns. In this study the younger sibling developed MDS with profound cytopenia, whilst the older remained in a pre-malignant stage presenting with mild dysmegakaryopoiesis. The authors revealed the twins had differing methylation patterns at the promoter region of GATA2, with the younger sibling having a more densely methylated GATA2 promotor region, which is likely to result in a lower GATA2 expression. As previously described additional mutations or cytogenetic abnormalities are necessary for malignant transformation of haematopoietic cells, the two sisters also had different somatic mutations between them, the number and order in which they were acquired could also contribute to the variation between the two phenotypes which are derived from the same GATA2 mutation (Kim *et al.*, 2019). Taken together, these cases show that allele-specific changes to GATA2 expression driven by epigenetic mechanism could be one reasonable explanation for the varied clinical presentations among GATA2 deficient patients.

1.3.4.2.4 Dominant negative mutant protein

GATA2 mutants show variable loss of transactivation capacity at known GATA2-responsive elements such as LYL1 and CSF1R promoters and RUNX1 and CD34 enhancers (Hahn *et al.*, 2011). This altered transactivation has been shown to generally correlate with reduced or diminished DNA binding activity as key residues responsible for recognising the GATA2 consensus WGATAR DNA motif are commonly mutated in GATA2 deficient patients (T354M, T355del, R361L/C and R396Q)(Chong *et al.*, 2018).

GATA2 interaction and co-activation with PU.1 has been investigated and has been shown to differ between mutants in a non-DNA-binding dependent manner. The mutations T354M and C373R were found to bind PU.1 more tightly compared to wildtype GATA2 and is thought that this enhancement may prevent GATA2 and PU.1 in carrying out their normal roles in myelopoiesis. How GATA2 mutants' effect other binding partners such as CEBPA has yet to be investigated. However, perturbed interactions with key haematopoietic transcription factors could result in the dysregulation of normal cellular processes such as proliferation, differentiation and survival. The variable degree of DNA-binding activity, transactivation and changes to protein-protein interactions documented throughout the literature between the different mutation types could also contribute to the diverse clinical outcomes featured in GATA2-related disease (Chong *et al.*, 2018).

1.3.4.2.5 Evolution of GATA2 deficiency towards haematological malignancy

Patients with GATA2 deficiency frequently develop MDS at a median age of onset of 20 years, significantly younger than patients with idiopathic MDS occurring around 60 years. These immunodeficient patients have a lifetime risk of MDS of ~90% with a large proportion being diagnosed with MDS at disease presentation (30-50%) (Collin, Dickinson and Bigley, 2015b). Furthermore, a high risk of disease evolution to AML or CMML have also been documented (Wlodarski, Collin and Horwitz, 2017; Donadieu *et al.*, 2018), along with two cases with acute lymphoblastic leukaemia (ALL) (Koegel *et al.*, 2016)(Donadieu *et al.*, 2018). A recent account of rapid progression to AML in a 25 year old patient in a 4 month period has been described (McReynolds, Zhang, *et al.*, 2019). Many acquired somatic mutations have been associated with myeloid transformation. ASXL1 mutations were the first to be associated with GATA2 deficiency in earlier reports (Bödör *et al.*, 2012; West *et al.*, 2014), followed by mutations in the RAS pathway and mutations in MDS/AML related genes (Wang *et al.*, 2015). Aberrant cytogenetics have also been implicated in disease progression with monosomy 7 and trisomy 8 recurring most frequently in patients (41% and 15%)(Chou *et al.*, 2010; West *et al.*, 2014; Wlodarski, Collin and Horwitz, 2017; McReynolds, Yang, *et al.*, 2019). Further genetic analysis of GATA2 deficiency patients are vital in understanding key somatic mutations involved in the transformation of the disease, which could also inform on treatment options.

On observations of mutational status alone GATA2 related MDS has been shown to have a slightly reduced overall survival (OS) when compared with wild-type GATA2 groups (73% vrs 84%, $P<0.05$) (5-year survival rate). However, when looking at MDS subtype and presence of monosomy 7 no differences in patients with GATA2 mutations and those with wild type were observed. Indeed, this finding should be interpreted lightly as the disease (Wlodarski *et al.*, 2016). In contrast a study composed of a larger cohort of GATA2 deficient patients with a

higher median age of MDS onset reported a poor survival outcome (Based on the three IPSS-R groups; 30%-high risk group, 80%- intermediate group and 100% low-risk group). It is important to note that there was a high prevalence of severe complications from GATA2 related phenotypes such as infections from immunodeficiency that were the main cause of deaths in patients over 40 years of age (Donadieu *et al.*, 2018). This could account for the discrepancies in overall survival between these two studies as the previous cohort were children with MDS and advanced disease.

1.3.4.2.3.1 ASXL1 gene

The ASXL1 gene is a member of the ASXL family, is comprised of 13 exons located on chromosome 20 band q11 and is involved in epigenetic gene regulation. Each member of the ASXL1 family share a common domain architecture; a highly conserved ASX homology (ASXH) domain, located at the N-terminal region and a plant homeodomain (PHD) finger, at the C-terminal region. The PHD domain is thought to be responsible for the binding of histones with specific modifications and in the recruitment of chromatin modulators and transcription factors (Sanchez and Zhou, 2011). The ASXH domain facilitates interactions with BRCA1 associated protein-1 (BAP1), a partner protein which is an essential component of the 'polycomb repressive deubiquinase complex (PR-DUB), in which it is responsible for the unbiquitination of monoubiquinated histone H2A at lysine 119 (H2AK119ub) - a modification that is catalysed by the polycomb repressive complex 1 (PRC1)(Scheuermann *et al.*, 2010). This PR-DUB complex also contains other ASXL1 family members that are responsible for its deubiquinating activity (Sahtoe *et al.*, 2016). Along with BAP1, ASXL1 also interacts with catalytic and scaffold subunits of PRC2; EZH2, EED and SUZ12, which promote trimethylation of H3 at lysine 27 (H3K27me3)(Inoue *et al.*, 2013). Therefore, ASXL1 is considered to act as an epigenetic scaffold during the regulation of several histone modifications such as H2AK119ub and H3K27me3(Asada *et al.*, 2018).

1.3.4.2.3.2 ASXL1 mutations in haematological malignancy

Mutations in ASXL1 are frequently found in myeloid neoplasms including MDS (14-23%), CMML (40%-49%) and AML (5%-17%) (Fujino and Kitamura, 2020) **(Figure 1.7 A).** ASXL1 mutations, along with DNMT3A and TET2, have also been identified as one of the most frequently mutated genes in Clonal haematopoiesis of indeterminate potential (CHIP), a premalignant condition associated with increased risk of subsequent haematological malignancy. ASXL1 mutations observed in patients with CHIP co-exist with others such; epigenetic regulators IDH2 and EZH2, splicing factors SRSF2 and U2AF1, signal transducing molecules JAK2, NRAS and SETBP1, transcription factor RUNX1 and components of the

cohesion complex STAG2. Conversely, it has been noted that ASXL1, DNMT3A, FLT3-ITD, WT1 and NPM1 mutations are exclusive in CHIP(Tyner *et al.*, 2018; Fujino and Kitamura, 2020). Collectively, the involvement of ASXL1 in myeloid malignancy and CHIP indicates it to be one of the earlier genetic events during myeloid transformation(Fujino and Kitamura, 2020).

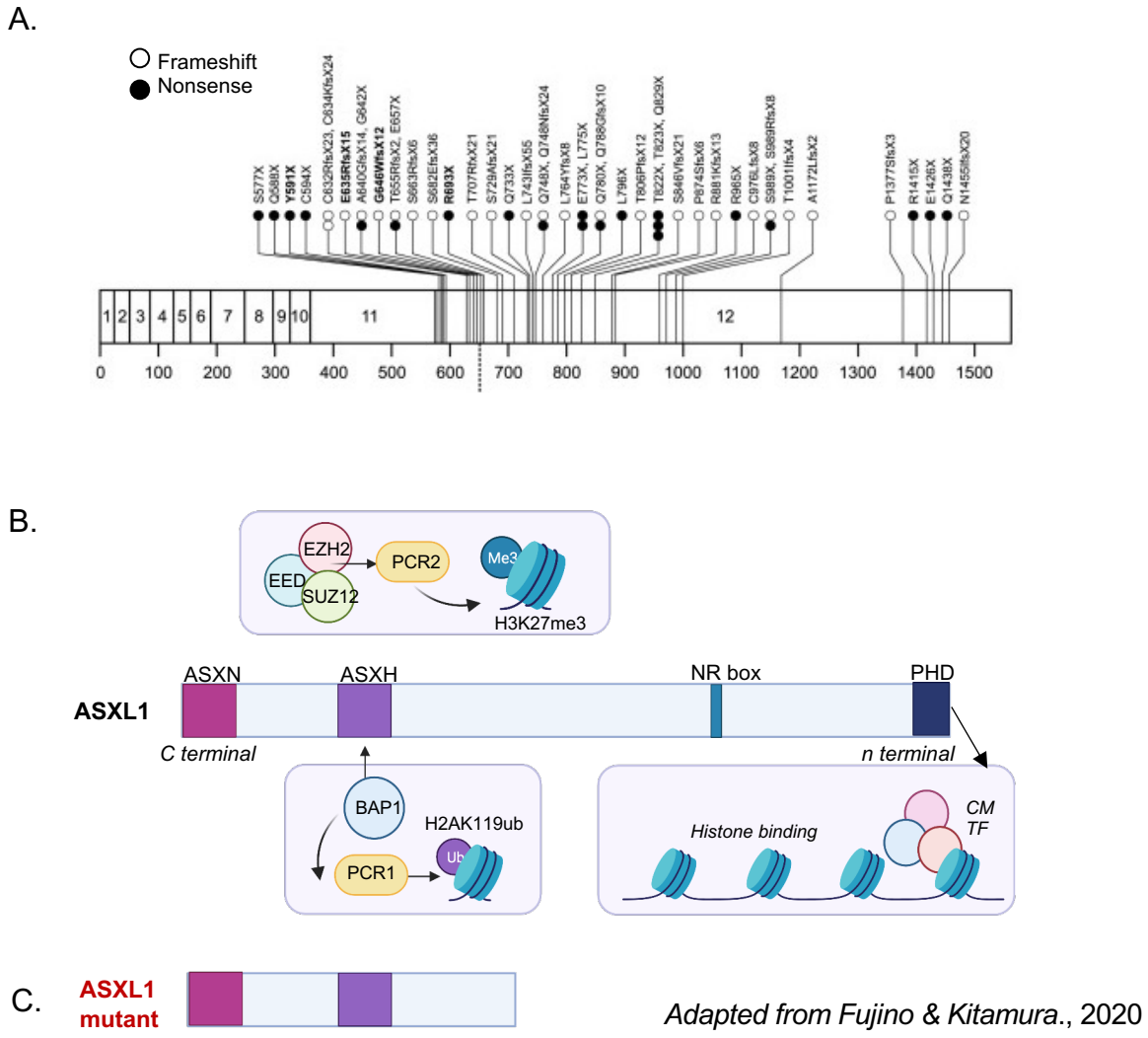


Figure 1.7: ASXL1 wild-type and mutant protein (A) Localization nonsense or frameshift mutations ASXL1 mutations documented in AML patients. These are concentrated in N-terminus of the last exon, generating a C-terminally truncated form of ASXL1 (Mutant ASXL1). Hotspot mutations are indicated in bold text. **(B)** ASXL1 wild-type protein, The ASXH domain facilitates interactions with BRCA1 associated protein-1 (BAP1), an essential component of the 'polycomb repressive deubiquinase (PR-DUB) complex responsible for the unbiquitination of monoubiquinated histone H2A at lysine 119 (H2AK119ub) - a modification that is catalysed by the polycomb repressive complex 1 (PRC1), ASXL1 interacts with catalytic and scaffold subunits of PRC2; EZH2, EED and SUZ12, which promote trimethylation of H3 at lysine 27 (H3K27me3). The plant homeodomain (PHD) is suggested to aid

binding of histones with specific modifications and aids recruitment of chromatin modulators and transcription factors. **(C)** The resulting mutant ASXL1 protein generated from mutations.

Mutations documented in ASXL1 are predominantly heterozygous frameshift or nonsense mutations located in the N-terminal region of the last exon. These mutations produce a C-terminally truncated protein adjacent to the PHD domain that escape nonsense-mediated decay mRNA decay (NMD) (Gelsi-Boyer *et al.*, 2009; Abdel-Wahab *et al.*, 2012) It has also been noted that ASXL1 mutations are also associated with a poorer prognosis (Schnittger *et al.*, 2013). The mechanism of how these ASXL1 mutations cause myeloid transformation is still yet to be fully defined. Nevertheless, consistent with the loss of function model, studies have demonstrated that ASXL1 mutants inhibits PRC2, leading to the upregulation of target genes in a dominant negative fashion. ASXL1 mutants involved in the ASXL1/BAP1 complex become hyperactivated, resulting in the promotion of aberrant myeloid differentiation in haematopoietic progenitors. Mechanistically this is mediated through upregulation of HOXA genes and IRF8. (Inoue *et al.*, 2013; Balasubramani *et al.*, 2015; Asada *et al.*, 2018) **(Figure 1.7 B)**. Collectively, these studies suggest that ASXL1 mutations do not simply cause a loss of function but may also involve a dominant negative mechanism.

1.3.4.2.3.3 ASXL1 in GATA2-related MDS/AML

Indeed, ASXL1 mutations have been documented to have a vital impact on haematopoiesis and the transformation to myeloid neoplasms. As previously described this transformation is also a common feature of GATA2 immunodeficiency. ASXL1 mutations have been documented as a common secondary event in GATA2-related myeloid malignancy and is thought to be a collaborating event in disease progression. The first cases were reported in a pedigree carrying the T354M GATA2 mutation of which two first degree cousins developed high risk MDS with monosomy 7. Upon further screening for additional recurrent mutations, identical somatic ASXL1 mutations were also observed (Bödör *et al.*, 2012). Following on from this initial discovery, West *et al* and colleagues conducted screening in a cohort of GATA2 deficient patients for ASXL1 mutations located in exons 12 and 13, previously associated with the transformation of MDS to AML. Heterozygous ASXL1 mutations were observed in 29% of patients, including 4/5 patients with proliferative CMML. Whilst a similar incidence of myeloid transformation was documented in patients with GATA2 deficiency in comparison to patients with ASXL1 mutations in the literature, interestingly, co-occurrence of germline GATA2 and somatic ASXL1 mutations were observed exclusively in females which were considerably younger and had a high incidence of transformation to CMML(West *et al.*, 2014). Moreover, a large proportion of these patients (46%) were observed to have unfavourable cytogenetics (monosomy 7, trisomy 8, monosomy 6) as observed by Bödör *et al*. However, a large

proportion of GATA2 deficient patients without somatic ASXL1 mutations also were reported to have aberrant cytogenetics (monosomy 7/ trisomy 8). Therefore, abnormal cytogenetics do not appear to be a specific feature of GATA2 deficient patients with somatic ASXL1 mutations. Since these initial papers, there is not much further information on the prevalence of somatic ASXL1 mutations among GATA2 deficient patients in the literature.

Further analysis the five-generation pedigree reported by Bodor *et al* (2012) was conducted by Seraihi *et al* in 2017 who reported a third cousin (IV-10) who in the previous study had been diagnosed with a T354M GATA2 germline mutation but remained an asymptomatic carrier had presented with recurrent minor infections, monocytopenia and neutropenia over a 3-year period which thereafter stabilised. After targeted deep sequencing of 33 genes frequently mutated in MDS/AML, the patient shared an identical ASXL1 mutation to that of its cousins previously described, though at a lower variant allele frequency (VAF) (12-6% over 6 years, compared with 35%). This low VAF of acquired somatic mutation ASXL,1 along with lack of progression suggests that co-occurrence of GATA2 and ASXL1 mutations alone are not enough to promote clonal expansion and leukaemia transformation (Al Seraihi *et al.*, 2018). Interestingly, whether transformation requires the addition of monosomy 7 as a tertiary event for the disease progression or whether monosomy 7 must be the secondary event which is vital for clonal expansion of the ASXL1 clone for disease progression is unclear. Indeed, the important question being asked throughout the literature- Does mutation order matter in disease progression? This has been highlighted for mutations in cells carrying mutations in both JAK2 and TET2 that are implicated in MPN, where the abundance of progenitor cells produced by double mutant HSCs with JAK2 as the initiating mutation was increased in contrast to double mutants HSCs with TET2 as the initiating mutation suggesting an intrinsic epigenetic mechanism impairing HSC output (Ortmann *et al.*, 2015). This highlights how two mutations can have distinct biological and molecular properties and is also a key question that needs to be investigated during GATA2-related malignancy. It is therefore vital, to screen GATA2 deficient patients in order to understand the genetic landscape to gain more knowledge of disease progression.

1.3.4.2.3 Modelling GATA2 immunodeficiency

Despite the large number of GATA2 mutations described throughout the literature, how GATA2 heterozygous mutations affect the haematopoietic system, cause multilineage cytopenia and transform into leukaemia is still not well understood.

1.3.4.2.3.1 Mouse models

Many mouse models have been employed to study GATA2 deficiency however, the majority of this work has been ineffective at recapitulating the human phenotype observed in patients. As discussed previously Gata2-homozygous KO mice were shown to be embryonic lethal due to a lack of haematopoietic development, whilst heterozygous mice revealed a reduction in the abundance and expansion of HSCs in the AGM region and though no defects were observed in YS, FL and BM HSCs from all haematopoietic sites revealed qualitative defects during serial or competitive transplantation assays (K.-W. W. Ling *et al.*, 2004). In line with this HSC defect at development, studies investigating *Gata2* haploinsufficiency during adult haematopoiesis, observed reduced abundance of HSCs which were functionally impaired in competitive transplantation assays. Additionally, these studies were able to capture the GMP defect observed in GATA2 immunodeficient patients observed in these mice as a loss GMP abundance and functionality (Rodrigues *et al.*, 2005; Rodrigues, Boyd, Fugazza, May, Yan Ping Guo, *et al.*, 2008). In support of these findings, recent work conducted in our own laboratory also employing the VaviCre mouse model to specifically delete one Gata2 allele observed a reduction in the abundance of cells in the GMP compartment along with reduction in the frequencies of Pre-GM, GMP, Pre-MegE, MkP, and Pre-CFU-E cells in young mice. Additionally, during aging these mice revealed a lymphoid defect specifically in the CLP compartment (Abdelfattah, 2020). Though some defects to haematopoiesis have been highlighted during these studies, they fail to recapitulate the hallmark cytopenias such as the loss of monocytes, B cells and NK cells that are well documented in patients with GATA2 deficiency.

Hypomorphic $G2^{fGN/fGN}$ **mice**

Research completed by Harada *et al* (2019) investigated how the reduced Gata2 expression affects adult haematopoiesis, to do this they employed their previously generated hypomorphic mutant mouse model ($G2^{fGN/fGN}$), which expressed reduced GATA2 expression to 20% of WT mice. This model escapes the early embryonic lethality observed with Gata2-KO mice, though authors previously reported it does not completely support urological organogenesis in some animals (Hoshino *et al.*, 2008). The authors document that quantitative GATA2 deficiency leads to minor haematological defects in young $G2^{fGN/fGN}$ mice as animals presented with thrombocytopenia persisted throughout life. Interestingly, they found that the incidence of leukocytosis increased with age and was associated with a worsening thrombocytopenia and anaemia. It is also important to note that whilst the majority of cases in these aging mice had incidences of leucocytosis, some $G2^{fGN/fGN}$ mice displayed WBC counts within the normal parameters (asymptomatic animals?). Further investigation into the peripheral blood of animals with overt leukocytosis, revealed a range of morphological

abnormalities such as; anisocytotic erythrocytes, hypogranular platelets, nuclear fragmentation and blast-like cells (typical of MDS) in comparison to animals without leukocytosis which displayed normal morphology. Hypermorphic expression in young mice also lead to impaired HSC function as self-renewal activity was reduced and progenitor cell differentiation skewed towards the granulomonocytic lineage. Mast cell development was impaired (rare in skin and absent in peritoneal lavage fluid), in part due to the skew in differentiation towards granulocyte-monocytic lineage. Moreover, myeloid cells that had the differentiation capacity toward the granulomonocytic lineage, were increased in animals from a young age, observed by the increase in GMPs within the BM which had a significant increase of monocyte colony-forming efficiency. This is contrary to previous studies reporting a significant reduction in GMPs in Gata2 haploinsufficient mice (Rodrigues, Boyd, Fugazza, May, Yan Ping Guo, *et al.*, 2008) which is also in line with observations of the BM of GATA2 immunodeficient patients (Bigley *et al.*, 2011). However, this could be attributed to the severe reduction in GATA2 expression in these animals, resulting in a CMP bias towards GMP. In further support of this argument during a cohort study mentioned in the paper the authors observed a survival rate of 70% (n=25) in $G2^{fGN/fGN}$ mice after ~16 months in comparison to a 100% survival rate of $G2^{+/fGN}$ (n=39) supporting the notion that the level of GATA2 expression could account for the differences in phenotypes. Mice also showed increased expression of transcripts encoding cytokine receptors vital for granulocyte and monocyte differentiation such as macrophage colony- stimulating factor (Csf1r) and interleukin- 6 receptor α- chain precursor (Il6ra) in GMPs. This was further confirmed by the over expression of macrophage colony-stimulating (M-CSF) and IL-6 receptors by flow cytometry. Upon exposure to lipopolysaccharide (LPS) which is known to trigger pro-inflammatory cytokine release, young mice developed severe granulomonocytosis along with a substantial decreased in B-lymphocytes resembling multilineage defects observed in human CMMoL. It would be interesting to see whether these animals had acquired any additional mutations during the disease progression such as ASXL1 or cytogenetics aberrations. Taken together, this study suggests that hypomorphic expression of GATA2 predisposes aging mice to MPN or MDS and that GATA2 deficits in $G2^{fGN/fGN}$ young mice cause GMPs to become hyperreactive to inflammatory stimuli, which can subsequently provoke disease during inflammatory stress.

Patient +9.5kb enhancer mutations

Though most GATA2 immunodeficiency mutations occur in the GATA2 DNA binding ZF domains, cases of patients carrying mutations in the +9.5kb enhancer, causing deletion of the E-box or carry single nucleotide Ets mutations resulting in disruption of the GGAW motif essential for DNA binding have been described. Both of these types of motif mutations are reported to have reduced GATA2 expression (Kirby D. Johnson *et al.*, 2012; Hsu *et al.*, 2013b).

Work by the Bresnick group have looked into the functional consequences of +9.5 enhancer mutations using specific mouse models.

A mouse strain was generated by targeted heterozygous deletion of the E-box and GATA motif in the GATA2 +9.5kb enhancer using homologous recombination in order to model the 28bp deletion observed in a patient presenting with Mycobacterial infection (MonoMac syndrome) and myelodysplasia. In comparison to their homozygous counterpart that were embryonic lethal at E13.5-14.5 due to severe haemorrhaging and variable edema, heterozygous embryos were viable and exhibited defects in definitive but not primitive fetal liver HSPC activity, along with a reduced expression of Gata2 and its target genes (Kirby D. Johnson *et al.*, 2012).

Ets mutation (1017+572C>T)

More recent work employing a mouse strain harbouring the heterozygous disease Ets mutation (1017+572C>T) observed in GATA2 immunodeficiency patients, observed though embryos were morphologically normal, HSC emergence, LT-repopulating fetal liver HSCs and Gata2 expression were 50% lower than WT animals. To assess whether this defect continued throughout embryogenesis immunophenotypic HSCs and MPPs were analysed from E15.5 FL and revealed HSC were unaffected however interestingly MPPs were around 2-fold higher. On assessment of adult haematopoiesis in the (-/-) mice parameters such as peripheral white blood count (WBC), red blood count (RBC) and platelet numbers were comparable to the WT mice. This lead to the conclusion that in steady state, this predisposition mutation in a murine disease model, generated measurable though functionally “silent” molecular alterations (Soukup *et al.*, 2019)- further supported by RNA-seq data generated from (-/-) mutant HSPCs cells observing no changes to the HSPC transcriptome compared with wildtype. Authors suggested that to observe the emergence of a pathogenic phenotype this could require an increase in demand on the haematopoietic system which would trigger quiescent HSCs to proliferate and generate HSPC regeneration. Thus myeloablation-induced stress was initiated using 5-FU in (-/-) mice and revealed abolished GATA-2 induction and reduced haematopoietic regeneration of the BM. To assess whether the myeloablation-induced stressed had affected gene expression changes, RNA-seq was repeated in HSPCs post treatment and revealed 2974 differentially expressed (FDR 0.05 and minimum fold-change of 2) genes between WT and (-/-) mutant mice, showing a clear change to the HSPC transcriptome upon secondary insult. The most deregulated genes observed were those involved in cell cycle regulation specifically in G2/M checkpoints, cellular proliferation (Myc and E2f targets), genes involved in mitotic spindle regulation and mTORC1 signalling (Soukup *et al.*, 2019).

1.3.4.2.3.2 Human models

Studies using primary HSPCs from patients with GATA2 deficiency syndromes would be the most ideal method for investigation disease pathology and progression however, this remains challenging due to the marked reduction of cellularity in the BM prior to the development of overt MDS or AML (Ganapathi *et al.*, 2015). Nevertheless, whole transcriptome sequencing of HSPCs from eight MDS patients carrying germline GATA2 mutations was carried out in a recent publication which highlighted distinct alterations in the transcriptome within specific lineages (Wu *et al.*, 2020). The HSC compartment in these patients highlighted dysregulated gene expression in relation to quiescence, cell cycle and apoptosis, along with increased expression of erythroid/megakaryocytic priming genes and decreased lymphoid priming genes. In line with patient immunophenotyping in the literature, a reduction in lymphoid/myeloid progenitors suggested to be a result of the altered HSC transcriptome prior to commitment. Moreover, strikingly the most dysregulated genes in both HSCs and HSPCS were those enriched in immune responses. Analysis of cells with chromosomal defects such as those with trisomy 8 and complex cytogenetics were observed to have downregulation of DNA repair genes and defects in genes related to multilineage differentiation and cell cycle respectively. (Wu *et al.*, 2020). This research has provided key data for determining molecular mechanisms of GATA2 immunodeficiency which require further investigation. Of particular interest was the skewed immune and erythroid gene expression from these patients.

Models employing induced pluripotent stem cell (iPSC) have been utilised to overcome the lack of primary material in a variety of different inherited bone marrow failure syndromes including GATA2 deficiency (Jung, Dunbar and Winkler, 2015). To date only two groups have reported successful iPS cell reprogramming from patients with GATA2 mutations and have investigated how these mutants effect the development of haematopoietic lineages (Kotini *et al.*, 2017; Jung *et al.*, 2018).

Kotini *et al* (2017) were able to generate a collection of iPS cell lines from patients with myeloid malignancy, encompassing a range of different disease stages. One of the four patients reported was diagnosed with high risk MDS and was identified to carry a germline GATA2 mutation T357N (VAF 0.50), along with a translocation (accounting for 99% of cell by FISH) and three recurrent gene mutation, of which the last was a second GATA2 mutation 390delK (VAF 0.27). Two human iPS cell lines were generated; the first to mirror the pre-leukemic state involving the initial germline GATA2 mutation and the second representing the disease state at collection (high-risk MDS), composing all detected mutations. Characterisation of these two lines during *in vitro* haematopoietic differentiation captured phenotypes of different severity

when compared to control lines. Both lines gave rise to comparable numbers of HPSC (CD34+) at early stages of cultures (D8), suggesting no early developmental defects in mesoderm formation or haematopoietic lineage specification. The high risk MDS line produced an increased abundance of mature HPCs (CD45+) which was delayed and at an overall reduced efficiency, in contrast to the pre-leukemic and control line which generated HPCs that matured beyond the progenitor stage observed from the loss of CD34+ expression in over 90% of the population. Megakaryocyte progenitors ($CD41a^{+}/CD45^{-}$) usually emerging on day 8 of culture in control cells was significantly reduced in the pre-leukemic line and completely abolished in the high risk MDS line. Furthermore, reductions in clonogenicity were observed, with BFU-E and CFU-GEMM being mainly affected in preleukemic lines in comparison to the almost complete abolishment in high-risk MDS lines. Finally, morphological analysis of HPCs at day 14 of differentiation or the more mature cells from CFC assays revealed mild dysplastic changes restricted to the erythroid line in the preleukemic line, whereas these changes were more widespread affecting all lineages in high-risk MDS lines. Though multilineage dysplasia phenotype in the high-risk MDS iPS cell line seen during this research recapitulated the MDS disease observed in patients, the pre-leukemic line which serves as a model for the initial events during patients with GATA2 deficiency showed no defects. Mutation of the second GATA2 allele (also reported in some patients) in this line using CRISP/Cas9 gene editing suggested that the additional loss of function of GATA2 contributes to disease progression, though also points out that additional events are needed to progress to a more aggressive disease as observed in the deletion of 7q in the preleukemic line causing abolishment of HPCs.

A second study conducted by (Jung *et al.*, 2018) successfully reprogrammed iPS cell lines derived from GATA2 deficient patients with mutations either in the ZF2 domain (R337X and R361H) or within intron 5 (Int5) one of which was asymptomatic. In line with the previous research no major developmental defects were observed as; mesoderm precursors ($CD31^{-}CD43^{-}CD73^{-}KDR^{+}APLNR^{+}$ cells) at day 4, Hematoendothelial progenitors (HEPs) ($VE\text{-}Cadherin^{+}CD73^{-}CD43^{-}CD235a^{-}$) at day 9 and HSPCs ($CD34^{+}CD45^{+}$) at day 16, were all comparable between fibroblast (FB)-derived and MSC derived control iPS lines and lines containing ZF mutations and Intron 5 mutations. Nevertheless, intron 5 mutant lines produce more mature HPCs marked by CD45+ expression and peripheral blood mononuclear cell (PBMC)- derived iPS lines did reveal an increase in the CD34+CD45+ commitment between control and Int5F iPS lines. Taking into consideration the distinct lineage defects observed in GATA2 deficient patients, no differences were observed in the total number of colonies during CFU assays. However, the number of CFU-GM colonies were significantly reduced in all mutation lines along with an increased in CFU-M colonies. Furthermore, though GATA2 has

been reported to be critical in the ETH transition (Kang *et al.*, 2018), functionality of HEPs derived from an *in vitro* HEP-OP9 co-culture method, established by Choi *et al.,* (2012) showed a modest decrease in haematopoietic potential observed by assessing the CD43+/CD45+ fractions. A common presentation of GATA2 deficiency is severe NK lymphopenia, which often results in increased susceptibility of viral infections. Surprisingly mutant HSPCs exhibited no defects in differentiating towards the NK lineages defined as $CD45^{+}CD3^{-}CD56^{+}$ and even GATA2 KO iPS lines showed unperturbed NK differentiation. It has been documented that patients with GATA2 mutations have residual NK cells in the PB which decrease over time and during disease progression. It is therefore suggested that the NK differentiation is not impaired in these patients, though the production of NK cells decline as HSPC abundance declines in the BM (Corat *et al.*, 2017; Schlums *et al.*, 2017). Interpretation the results from this study should be carried out with care as the authors noted overall a low and variable efficiency of the control lines.

Taken together, these studies have investigated the effect of two of the four classes of GATA2 germline mutations; missense (R357N, R361H and R337X) and mutations resulting in aberrant mRNA splicing (Int5) during haematopoietic development using patient derived hiPS modelling. Collaboratively, this research suggests that during steady state, GATA2 mutations result in very subtle defects to the haematopoietic system and that extrinsic factors such as epigenetic or environmental factors may play a substantial role in pathogenicity. This is supported by the fact that new-borns with GATA2 mutations are born with functional haematopoietic systems and that disease usually manifests within the first decade of life, along with the observations in these studies that activation of a stressor such as myeloablative stress or proinflammatory stimulation result in the development of disease. The severe immunological challenge that GATA2 deficient patients face throughout life are well documented and these insults are not represented in normal mouse and iPS models unless artificially stimulated. How an inflammation rich environment interacts and or changes mutant cell survival and function is becoming a hot topic in myeloid malignancy as observed in CHIP studies. This could also be key in understanding GATA2-related myeloid malignancies. This research also presents for the first time, the opportunity to investigate the differences at the molecular level of two very different stages of GATA2 deficient disease from a single patient, the initial GATA2 mutation in comparison to high-risk MDS from acquisition of and cytogenetic abnormalities.

Unfortunately, both studies failed to recapitulate the specific lineage-differentiation abnormalities observed in patients carrying GATA2 mutations though, Kotini *et al* (2018) was able to recapitulate the multilineage dysplasia similar to what is seen in the clinic in the high-

risk MDS line. This could imply that the role of GATA2 at embryonic haematopoietic pathways might be distinct from those in adult HSPC. However, human iPS cell studies are limited by the haematopoietic differentiation culture systems that are currently available which follow a linear progression towards embryonic HSPCs. This also highlights a limitation of iPS models as they are only able to reflect primitive haematopoiesis. The assessment of self-renewal and multilineage engraftment potential by iPSC-derived HSPCs from these lines could be vital for understanding human patient disease in the context of GATA2 deficiency yet, current haematopoietic differentiation methods are unable to produce engraftable HSPC without enforced expression of various TF (Sugimura *et al.*, 2017).

1.4 Aims and objectives of thesis

Germline heterozygus mutations in the GATA2 gene gives rise to an immnodeficiency syndrome characterised by cytopenias, severe infections and a predisposition towards hematological maligancies such as myeloid dysplasia and leukaemia (Dickinson *et al.*, 2014; Spinner *et al.*, 2014). To date there have been over 350 genomic variations of GATA2 gene which result in many different phenotypes. Some mutations cause a quantitative GATA2 deficit whilst others lead to the production of mutant GATA2 proteins with altered expression (Shimizu and Yamamoto, 2020). Acquired somatic mutations in ASXL1 and cytogenetic abnormalities such as monosomy 7 and trisomy 8 are commonly found in patients that have progressed towards MDS/AML (Bödör *et al.*, 2012; West *et al.*, 2014; Bluteau *et al.*, 2017). Whilst there has been much research over the years investigating the loss of GATA2 in a variety of mouse and human models, many have failed recapitulate key aspects of the disease observed in patients such as multilineage disfunction and progression of disease. Therefore, the purpose of this thesis is to investigate a more specific model of GATA2 immunodeficiency and its progression through the following aims:

1. Evaluate the potential of human iPSC lines for in their use in downstream assays as a model of disease progression in terms of their; pluripotency status, genetic stability and their ability to differentiate efficiently into the haematopoietic lineage.

2. Engineer human iPSC lines carrying patient specific GATA2 SNPs along with lines carrying additional secondary mutations in the ASXL1 gene observed in 30% of GATA2 patients using CRISPR/Cas9 gene editing and CRISPR Cas9 base editors.

3. Assess the potential role of inflammation as a driver of disease progression at the GATA2 immunodeficient stage of disease through investigating shRNA-mediated knockdown of GATA2 in CB $CD34^+$ HSPCs in the presence of LPS stimulation and Gata2 heterozygosity using the Vav-iCre conditional mouse model during chronic pIpC stimulation.

Chapter 2

Materials and Methods

2.1 Cell lines

All cell lines used in this research were cultured at 37°C with 5% CO_2 and 20% O_2 and tested negative for Mycoplasma.

The murine OP9 cell line (ATCC® CRL-2749™) was purchased from ATCC strain (C57BL/6 x C3H) F2 -op/op. This line was established from new-born op/op mouse calvaria and do not produced functional macrophage colony-stimulating factor (M-CSF). Cells were maintained on gelatine-coated (0.1%) tissue culture flasks using α-MEM media supplemented with 20% FBS (non- heat activated) (Gibco).

The murine MS5 cell line was kindly gifted to our lab by Dr Fernando Dos Anjos (Cardiff University). This stromal cell line was established by irradiation of the adherent cells in long-term bone marrow cultures derived from C3H/HeNSlc strain mice. Cells were maintained in Iscove Modified Dulbecco Media (IMDM) (Gibco) supplemented with 10% FBS and Penicillin-Steptomycin (100U/mL; Gibco).

The HEK293T (Cell Biolabs) cell line is a derivative of human embryonic kidney 293 cell (HEK293T), modified to contain the SV40 T antigen. When used in conjunction with vectors carrying the SV40 region of replication it facilitates transient replication at a high copy number. HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS (heat-inactivated; Gibco) and 2mM L-glutamine.

The THP-1 cell line was kindly gifted to our lab by Dr Dipak Ramji (Cardiff University). This line is a human monocytic cell line derived from the peripheral blood of a 1-year old male with Acute Monocytic Leukaemia, FAB M5. THP-1 cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 Medium (Gibco) supplemented with 10% FBS (heat-inactivated) and 2mM L-glutamine.

The K562 cell line was kindly gifted to our lab by Dr Richard Darley at Cardiff University. This line is a human erythroleukaemia cell line derived by pleural effusion from a 53-year old female with chronic myelogenous leukaemia in terminal blast crises. K562 cells were maintained in RPMI 1640 Medium supplemented with 10% FBS (heat-inactivated) and 2mM L-glutamine.

The Human ES cell line H7 (WA07) was kindly gifted to our lab by Professor Meng Li (Cardiff University). This cell line was derived from a cleavage stage human embryo produced by *in vitro* fertilisation (IVF). The embryo was cultured to the blastocyst stage and cells from the

inner cell mass isolated resulting in ESC cell line H7 with a normal XY karyotype (Thomson, 1998).

The human iPS cell line KOLF2 was kindly gifted to our lab by Dr Branko Latinkic (Cardiff University). This line was derived from fibroblast cells from a 55-year old male using the CytoTune-iPS Sendai Reprogramming system containing four SeV based reprogramming vectors, each expressing one of the four Yamanaka factors (Oct4, Sox2, Klf4 and c-Myc).

The human Episomal (EPI) iPS cell line was kindly gifted to our lab from Dr Ashleigh Boyd (University College London; originally purchased from ThermoFisher). This line was derived from CD34+ cord blood cells using a three-plasmid seven factor EBNA-based episomal system (SOKMNLT; SOX2, OCT4 (POU5F1), KLF4, MYC, NANOG, LIN28, and SV40L T antigen)(Burridge *et al.*, 2011).

2.2 Human iPS cell culture

2.2.1 Essential 8 culture system

The human episomal iPS line was maintained under feeder-free conditions using Essential 8™(E8) Medium (ThermoFisher), containing E8 Basal Medium supplemented with E8 supplement (50X). Briefly, culture vessels were coated with Vitronectin (VTN-N) (ThermoFisher) diluted in Dulbecco's PBS (DPBS) without Calcium and Magnesium (ThermoFisher) at a working concentration of 5µg/mL and incubated at room temperature for one hour. Diluted VTN-N was aspirated, complete E8 medium, cells passaged directly onto VTN coated culture dishes at seeding density appropriate to the surface area of the culture vessel and incubated at 37^{o}C with 5% CO_2. Human iPS cells were passaged every 2-4 days once 70% confluency had been reached. Confluent cell vessels were dissociated into small clusters using 1X Versene solution (Thermofisher); Versene solution was added to the cells, incubated at 37°C for 4-5', aspirated and E8 medium added to wells to break up colonies into clusters. Clusters were seeded into pre-coated VTN-N culture vessels containing E8 medium supplemented with RevitaCell™ (ThermoFisher) to a 1X final concentration to improve cell survival in the first 24 hrs.

2.2.2 StemFlex culture system

Alternatively, cells used for CRISPR/CAS9 gene editing or sorting experiments were transitioned onto StemFlex™ (SF) Medium (ThermoFisher) containing SF Basal Medium supplemented with SF 10X supplement, and cultured on recombinant Human Laminin-521 (ThermoFisher) coated culture vessels to maximise cell performance during single cell culture. Briefly, culture vessels were coated with rhLaminin-521 at a working concentration of 0.5-2.0 µg/cm^2 diluted in DPBS (with Calcium and Magnesium) and incubated at 37°C for two hours. Cells were passaged into a single cell suspension using prewarmed TrypLE™ (ThermoFisher) , incubated at 37°C, 5% CO_2 for 5' followed by gentle pipetting.

2.3 Cryopreservation of cell lines

Cell lines were frozen down with 70% media, 20% FBS and 10% DMSO. Human iPS cell lines were frozen down using PSC Cryopreservation Kit (ThermoFisher) or alternatively 90% KnockOut Serum Replacement (ThermoFisher) and 10% DMSO. Primary Cord Blood cells were frozen down with, 75% IMDM medium, 15% FCS and 10%DMSO. All lines were collected from respective cultures, pelleted by centrifugation, resuspended in ice-cold freezing medium, transferred to cryovials (Camlab limited) and placed in a Mr. Frosty freezing container (Thermofisher) filled with 100% isopropanol (IPA) (Sigma). Freezing containers were placed in a -80°C freezer allowing cells to cool at a -1°C/minute rate. After a minimum of 4hrs,

cryovials were removed and stored at -80°C for short term storage or alternatively placed in liquid nitrogen (-175°C) for permanent or long-term storage.

2.4 RNA extraction and gene expression analysis

Samples from cultured cells were washed with DPBS, pelleted by centrifugation for 5' at 300xg, aspirated and the pellet stored at -80°C prior to RNA extraction. Extraction of RNA was carried out using the RNeasy Plus Micro kit (Qiagen) according to manufacturer's instructions. Briefly, cells were lysed with Buffer RLT, passed into a gDNA Eliminator spin column and centrifuged to remove genomic DNA. 70% ethanol was added to the flow through to adjust RNA binding conditions and transferred into a RNeasy MiniElute spin column and centrifuged. The column was washed with Buffer RW1 and Buffer RPE followed by 80% ethanol before finally eluting total RNA with RNase-free dH_20. RNA was subsequently stored at -80°C until use. cDNA synthesis was carried out using the QuantiTect Reverse Transcription Kit (Qiagen). Real time quantitative PCR (RT-qPCR) was carried out in the QuantStudioâ 7 Flex Real-Time PCR System (Applied biosystems) using either the Taqman (Applied biosystems) or Sybr green (PCR Biosystems) assay. Relative quantitation was calculated by comparing the Ct value of the target gene against the house keeping gene (GAPDH) within a sample using the formula: 2^{DDCT}. Primer information is listed in **Table 2.1**.

Table 2.1: RT-qPCR primers.

TaqMan Assay		
Primer	**ID name**	**Manufacturer**
GATA2	Hs00231119_m1	Applied Biosystems
ASXL1	Hs00898213_m1	
GAPDH	Hs02786624_g1	

2.5 Immunofluorescence

Cell lines were cultured on top of glass slides and maintained to confluency and fixed with 4% paraformaldehyde (ThermoFisher) for 10'. Slides were washed three times with 1X DPBS, permeabilised in 0.5% triton (in 1X DPBS) for 10' at RT, blocked in 1% BSA, 0.05% Tween in DPBS for 10' at RT and washed three times in 1X DPBS. Slides were incubated with primary antibody in 1% BSA in 1XDPBS (1:200) overnight at 4°C, washed three times and incubated with secondary antibodies (1:200) for 1hr at RT. Slides were washed three times and incubated with DAPI (1µg/mL) for 5' at RT in the dark. Slides were washed three times and mounted to slides using Mowiol 4-88 (Sigma) and imaged using a Leica Time-Lapse Microscope. See **Table 2.11** for full antibody information.

2.6 Chromosome quality assessment

Chromosome Harvesting was prepared as described by (Moralli *et al.*, 2011) with slight modifications. 1-2x10^6 human PS cells were seeded into a VTN coated 6 well plate in the presence of RevitaCell™ and cultured overnight. The following day cells were treated with the mitotic synchronizing agent Nacodazole (Sigma) at 0.1mg/ml and incubated for 16hrs. Cells were washed with DPBS and harvested by Tryple™ to generate a single cell suspension followed by centrifugation at 365xg for 10'. Cells were gently resuspended in buffered hypotonic solution (0.4% KCL with HEPES, pH 7.4) and incubated for 30' at 37°C. Cells were centrifuged, re-suspended in cold Carnoy's fixative solution (3:1 (v/v) methanol/glacial acetic acid), incubated for 30' at room temperature, centrifuged, washed in fixative and incubated for a further 10'. Following a final spin, cells were resuspended in a small volume of fixative. Finally, the cell suspension was dropped onto pre-chilled Polysine Adhesion glass slides (Thermofisher) from a height of 30 cm and subsequently placed on a heat block (60°C) for 5' to promote bursting of the nuclear membrane and the even spread of chromosomes. To analyse chromosome number, slides were covered with 10µg/mL 4′, 6-diamidino-2-phenylindole (DAPI; ThermoFisher) solution, incubated in the dark for 10', washed with DPBS and mounted using Mowiol 4-88 (Sigma). Mitotic chromosome counts were imaged using a Leica Time-Lapse Microscope and subsequently counted using ImageJ software.

2.7 Human iPSC differentiation

2.7.1 Human iPSC differentiation in OP9 co-culture

Overgrown OP9 cultures were prepared by seeding OP9 cells onto gelatinized (0.1%) T75cm^2 flasks and allowing them to further grow for an additional 2-4 days post confluency with continued feeding. Undifferentiated epi-hiPS cells were harvested with Versene and maintained in small clumps. Simultaneously, a small aliquot of cells was further dissociated to a single cell suspension and used for counting. OP9 culture medium was aspirated and human iPS cells were added to overgrown OP9 culture at a density of 1-1.5 x10^5 cells/6 well plate in 10mL of α-MEM supplemented with 10% FBS (non-heat inactivated), 100µM monothioglycerol (MTG) (Sigma), 50µg/mL Ascorbic acid (Sigma) and distributed by rocking the dish in a figure of 8. Human iPS/OP9 cell co-culture was incubated for 9 days at 37˚C with 5% CO_2, with the medium being replaced to 20mL on day 1, followed by a half-medium change on days 4, 6 and 8. hiPS/OP9 monolayers were dissociated into single cell suspension by sequential enzymatic digestion of collagenase IV (Sigma) for 30' at 37˚C, followed by 0.05% trypsin-0.5mMEDTA (ethylenediaminetetraacetic acid) for 15' at 37˚C at different time points of culture to follow the differentiation kinetics. Cells were washed twice with phosphate-buffered saline

(PBS) containing 2% FBS, filtered through a 100μM cell strainer and processed for CD34 enrichment.

2.7.1.1AutoMACS magnetic cell separation- CD34 Enrichment

Cell pellets from hEB/ OP9 co-culture were resuspended in 300ml of PBS 2% FBS containing 10ml human CD34 magnetic beads (Miltenyi Biotec) and incubated for 12' at 4°C while rotating. Cells were subsequently washed, filtered through a 30mm strainer (CellTrics) and fractionated in the AutoMACS using the "posselds" programme, used for separation of rare cells or to obtain high purity. The CD34+ cell population was then used for downstream cell sorting or immunophenotyping.

2.7.2 STEMdiff™ Haematopoietic Differentiation

Human iPS cells were differentiated to CD43+ primitive haematopoietic progenitor cells (HPC's) using the 12-day, two step differentiation protocol STEMdiff™ (Stemcell technologies). Human iPS cell lines, adapted to Matrigel (Corning) coated culture conditions (1:266 dilution in DMEM/F12 medium), were dissociated into small aggregates (~100 cells/ aggregate) using ReLeaSR (Stemcell technology), counted, 40 aggregates seeded into a 12 well plate and cultured overnight in E8 medium. The following day (Day 0) medium was aspirated and replaced with medium A (Basal Medium with 1x supplement A) to initiate mesoderm differentiation, followed by half a medium change at 48hrs (D2). On day 3 (72hrs) mesodermal cells were further differentiated into haematopoietic progenitors by replacing medium A with medium B (Basal Medium with 1x supplement B) with half a medium change on days 5, 7 and 9. At day 10 or 12 non-adherent cells were harvested from the culture and remaining adherent cells harvested with ACCUTASE™ (Stemcell Technologies). Haematopoietic differentiation was assessed through immunophenotyping.

2.7.3 Colony Forming Unit (CFU) Assay

Human clonogenic progenitor assays were performed by seeding sorted CD31+CD34+CD43+ populations either at day 14 (EB cultures) or day 9 (iPS-OP9 co-culture) into MethoCult™ SF H4636 (STEMCELL technologies). Briefly, cells were resuspended in 100μL IMDM, added to methylcellulose vortexed to distribute cells throughout media and plated using a 2mL syringe and 19G blunt ended needle. Cells were incubated at 37°C with 5% CO_2 and 20% O_2 and colonies counted between days 8-14 using an inverted microscope (Leica).

2.8 CRISPR/CAS9 editing

2.8.1 Forming CRISPR/CAS9 RNP complex

To generate heterozygous GATA2 mutations in human iPS cells, TrueGuide™ Synthetic gRNA was purchased from Invitrogen as a 2 piece crRNA:tracrRNA system, the crRNA was specifically designed to target GATA2 region. To generate the crRNA:tracrRNA duplex, TrueGuide™crRNA (100μM) and TrueGuide™tracrRNA (100μM) were mixed with 5X annealing buffer and nuclease-free water for a final crRNA:tracrRNA duplex concentration of 20μM and annealed in a thermocycler; 95˚C for 5', 95˚C to 78˚C with -2˚C/'' ramp rate, 78˚C for 10', 78˚C to 25˚C with -0.1˚C/'' ramp rate. 25˚C for 5'. To form the CAS9 Ribonucleoproteins (RNP) gRNA complex, 12.5μg GeneArt™ Platinum™ Cas9 Nuclease were premixed with 2.4μg crRNA:tracrRNA duplex incubated at room temperature for 10'.

2.8.2 Electroporation

Cells were transfected by electroporation using the Neon Transfection System (ThermoFisher). Briefly, confluent human EPI-iPS cells were dissociated as previously described using TrypLE™ to achieve a single cell suspension and $1x10^6$ cells resuspended in R solution. Cells were mixed with gRNA: CAS9 RNP complex along with the homology directed repair (HDR) template (3μL ws:100μM) and electroporated using a 100μl Neon tip using one 30-ms pulse at 1200 V. After electroporation, cells were seeded into a pre-coated rhLaminin 6-well plate containing prewarm complete SF medium. After incubation for 72hrs cells were stained for pluripotency surface marker TRA-1-60 and single cell sorted into 96 wells. TrueGuide™ Synthetic cRNA region of the guide and HDR template can be found in **Table 2.2**.

Table 2.2: Sequences for CRISPR/CAS9 editing of the GATA2 region T354.

Component	Sequence
T354 cRNA region gRNA	5' AG**A**C**G**ACAACCACCACC**TT**A 3'
HDR template (T354M)	5' CTGACGCTGCCTTGCCCTCCCAGTCGGCCGCCAGAAGAGCCGGCACCTGTTGTGC AAATTGTCAG**A**T**G**ACAACCACCACC**TT**GTGGCGCCGAAACGCCAACGGGGACCCTG TCTGCAACGCCTGTGGCCTCT 3'

2.8.3 DNA extraction and Genomic PCR

Genomic DNA (gDNA) was extracted from cultured cells using the ISOLATE II Genomic DNA Kit (Bioline). Briefly, cells were washed in DPBS, pelleted and resuspended in Lysis Buffer GL, Proteinase K solution and Lysis Buffer G3 and incubated at 70°C for 15'. 100% ethanol was added to each sample to adjust the DNA binding conditions, placed into an ISOLATE II Genomic DNA Spin Column and centrifuge for 1' at 11,000xg. Samples were washed with

Wash Buffer GW1 and GW2 and finally eluted in 70°C preheated nuclease free dH_20. DNA concentration was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific).

Genomic PCR was performed using Platinum™ Taq DNA Polymerase PCR kit (Thermo Scientific) according to the manufacturers protocol under the following conditions: Initial denaturing at 94°C for 1', 35 cycles of denaturing at 94°C for 15'', annealing was carried out at variable temperatures according to primer sequence **(Table 2.3)** for 30'', extension at 68°C for 1' and final extension at 68°C for 5'. Amplification was carried out using a T100 Thermal cycler (BioRad). Electrophoresis was carried out using a 2% [w/v] 1xTris-acetate-EDTA (TAE) agarose gel ran at 120V for 40'.

Table 2.3: Genomic PCR primers and corresponding annealing conditions.

Primer	Sequence	Annealing (°C)
GATA2 Fwd	5' TTAGCCCTCCTTGACTGAGC 3'	59
GATA2 Rev	5' GCACTCACATTGTGCAGCTT 3'	
ASXL1 Fwd	5' GCAGACAAACGGAACTCAACC 3'	60
ASXL1 Rev	5' AGTCCATGGTTGGTGCTACAT 3'	
RNF2 Fwd	5' GCAGACAAACGGAACTCAACC 3'	60
RNF2 Rev	5' AGTCCATGGTTGGTGCTACAT 3'	

2.8.4 Surveyor Assay

To detect single base substitutions incorporated into the gDNA from the CRISPR/Cas9 editing, the Surveyor Assay (Integrated DNA Technologies) was used. Extracted gDNA from individual clones were amplified as above using primers designed to the T354M region of the GATA2 gene (**Table 2.3**). Amplification was analysed from 2-5mL aliquots of each product by electrophoresis using a 1% [w/v] 1x TBE (89mM Tris-Borate [pH 8.3], 1mM EDTA) buffer agarose gel, ran at 120V for 40'. A 12mL digestion reaction was set up containing 100ng gDNA (8ml), 0.8ml 0.15M $MgCl_2$ solution, 1ml Surveyor Enhancer S and 1ml Surveyor Nuclease S and incubated at 42°C for 1hr to digest product. The reaction was inhibited by the addition of the stop solution, then cleavage assessed by gel electrophoresis in 1% [w/v] agarose gel in 1x TBE buffer. The GATA2 region of each clone analysed positive for the Surveyor assay was amplified as previously described and sent for DNA analysis at the Medical Research Council Protein Phosphorylation and Ubiquitylation (MRC PPU) Unit in Dundee Scotland.

2.9 CRISPR/Cas9 Base editing

To generate heterozygous point mutation at T354M and R361C in the GATA2 gene the base editing approach was employed outlined by (Komor, Badran and Liu, 2018).

2.9.1 DNA vectors and preparation

Plasmid phU6-gRNA was a gift from Charles Gersbach (Addgene) and the base editing plasmids pBK-YE1-BE3, phCMV BE3-Gam and phCMV BE4 max P2PA GFP were a gift from David Liu (Addgene). Plasmid details are located in **Figure 2.1(A) and Figure 2.2.** Plasmids were prepared by streaking bacteria stabs of corresponding plasmids onto 50mg/mL kanamycin (u6gRNA) or 100mg/mL ampicillin (base editor plasmids) resistant agar plates and incubating at 37°C overnight. Individual clones were picked from bacterial culture and inoculated in 250mL Luria broth (LB) (Sigma) with corresponding antibiotic overnight on a shaker at 37°C. A glycerol stock was made from 500mL inoculum and 500mL 50% glycerol solution, mixed and placed for long term storage in -80°C. The remaining culture was harvested, and plasmid DNA purified using the EndoFree Plasmid Maxi Kit (Qiagen) according to manufacturer's instructions. DNA was eluted in DNase free dH_20 at a concentration of >1mg/mL, aliquoted and stored at -20°C.

2.9.2 U6gRNA Cloning

Complimentary oligonucleotide sequences were designed for each of the guide sequences, containing mutations at the designed sites **(Table 2.4)** Oligo's were also designed to include a 5' G to enable transcription from the hU6 transcription, along with the sequence CCACC at the 5' end of the forward primer and AAAC at the 5' end of the reverse primer to complete the overhangs generated by the subsequent restriction digest.

Table 2.4: U6 gRNA oligo sequences.

Gene	Oligo	Sequence
GATA2	T354M Fw2	5' CACCGCAG**A**C**G**ACAACCACCACCTTA 3'
	T354M Rev2	5' AAACTAAGGTGGTGGTTGT**C**G**T**CTGC 3'
	R361C G1 Fw1	5' CACCGTTATGG**C**G**C**CGAAACGCCAA 3'
	R361C G1 Rev1	5' AAACTTGGCGTTTCG**GC**G**C**CATAAC 3'
	R361C G2 Fw2	5' CACCGTATGG**C**G**C**CGAAACGCCAAC 3'
	R361C G2 Rev2	5' AAACGTTGGCGTTTCG**GC**G**C**CATAC 3'
ASXL1	R1415* Fw1	5' CACCGTTACC**C**C**G**AGAGCCAGGGAA 3'
	R1415* Rev1	5' AAACTTCCCTGGCTCT**C**G**G**GGTAAC 3'
	Q1300* Fw1	5' CACCGGGC**C**A**A**GGGAAGAAGCTTTT 3'
	Q1300* Rev1	5' AAACAAAAGCTTCTTCCC**TT**G**C**CCC 3'
RNF2	Fw1	5' CACCGCAGGTAATGACTAAGATGAC 3'
	Rev1	5' AAAGCGTCATCTTAGTCATTACCTGC 3'

Complimentary Oligo's for T354M, R361Q-G1 and R361Q-G2 were annealed together and cloned into U6 gRNA expression vector (addgene #53188) using single type IIS restriction site

- BbsI which recognises the DNA sequence sense strand 5' GAAGAC$(N)_2$ 3' antisense strand 3' CTTCTG$(N)_6$ 5' **(Figure 2.1 A)** and T4 DNA ligase (New England BioLabs). The ligation reaction was transformed using one shot TOP10 Electrocomp cells (Thermofisher) according to manufacturer's instructions, inoculated in Luria broth (LB) overnight with Kanamycin. Liquid bacterial culture was isolated 24hrs later using Endofree Plasmid Maxi kit (Qiagen) according to manufacturer's instructions. Successful ligation was analysed by BbsI digestion of clone DNA, positive clones containing annealed oligo sequence result in the destruction of BbsI recognition sites, no digestion and therefore a circular plasmid. Samples were analysed by electrophoresis 2% [w/v] agarose gel in 1x TAE buffer and undigested samples sent for sequence for further confirmation (MRC PPU DNA Sequencing and Services) **(Figure 2.1 B)**

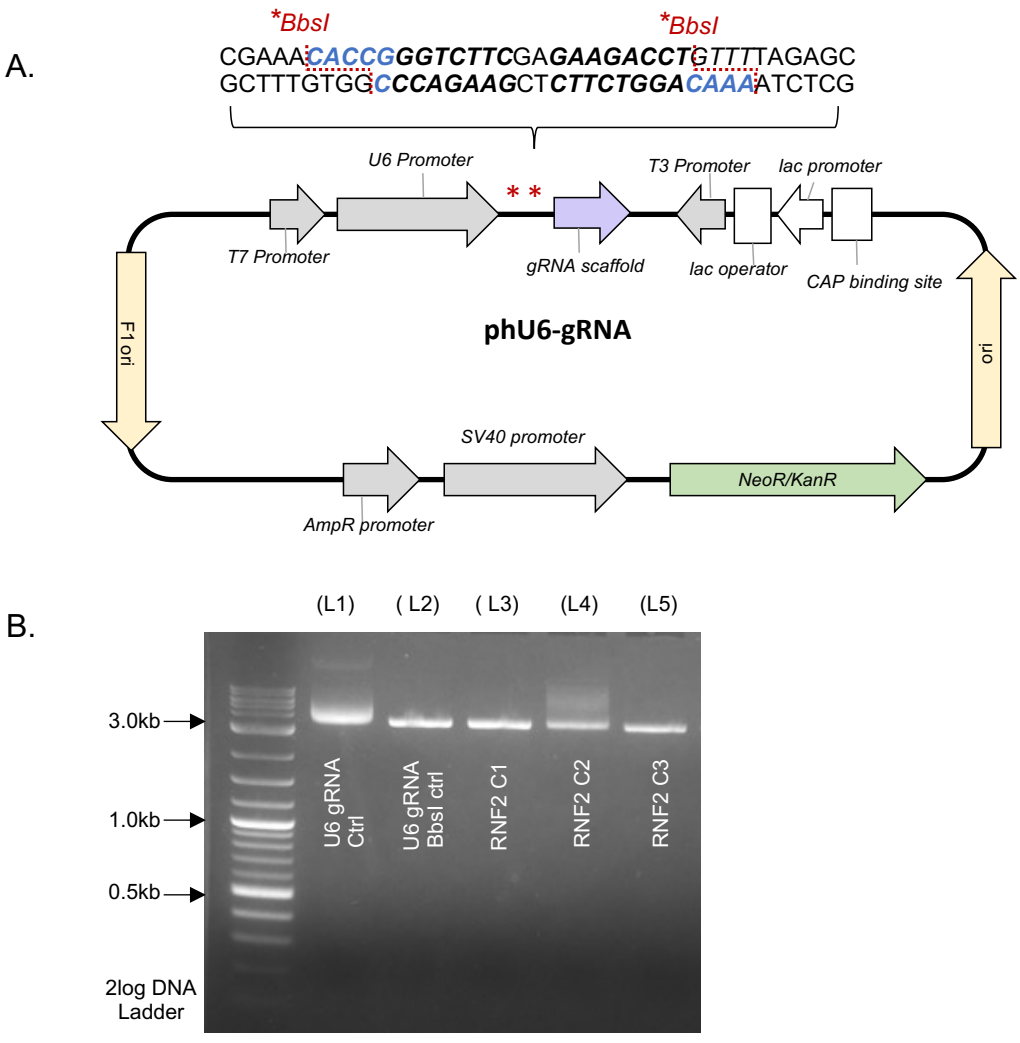


Figure 2.1: U6gRNA- cloning approach. (A) Plasmid map showing the type IIS restriction site BbsI which recognises the DNA sequence sense strand 5' GAAGAC$(N)_2$ 3' antisense strand 3' CTTCTG$(N)_6$ 5' (black bold) along with overhangs generated by the subsequent restriction digest (blue) which is found between the U6 promoter and the gRNA scaffold sequence. **(B)** Cloning results from RNF2 sequence cloning. Digested clones were resolved on a 2% [w/v] agarose gel in 1x TAE buffer against

undigested u6gRNA plasmid DNA (Lane 1), BbsI digested u6gRNA plasmid DNA with BbsI restriction sites intact, product size 3515bp (Lane 2). Lanes 3 and 5 show failed ligations, Bbs1 restriction sites intact resulting in successful linearization of the plasmid. RNF2 C2 (Lane 4) plasmid is has not been digested and circular plasmid is visualised by a smeared and diffuse band.

2.9.3 Base editing transfection

2.9.3.1 Transfection using Neon electroporation system

Human EPI-iPS cells were dissociated as previously described using TrypLE™ and $1x10^6$ cells resuspended in R solution. Cells were mixed with base editor (BE) plasmid (3.75µg) and u6 gRNA plasmid (1.25µg) in a 3:1 ratio and electroporated using a 100µL Neon tip using one 30-ms pulse at 1200 V. After electroporation, cells were seeded into a pre-coated rhLaminin 6-well plate containing prewarm complete SF medium supplemented with Revitacell (1x).

2.9.3.1 Transfection using Lipofectamine Stem Transfection reagent

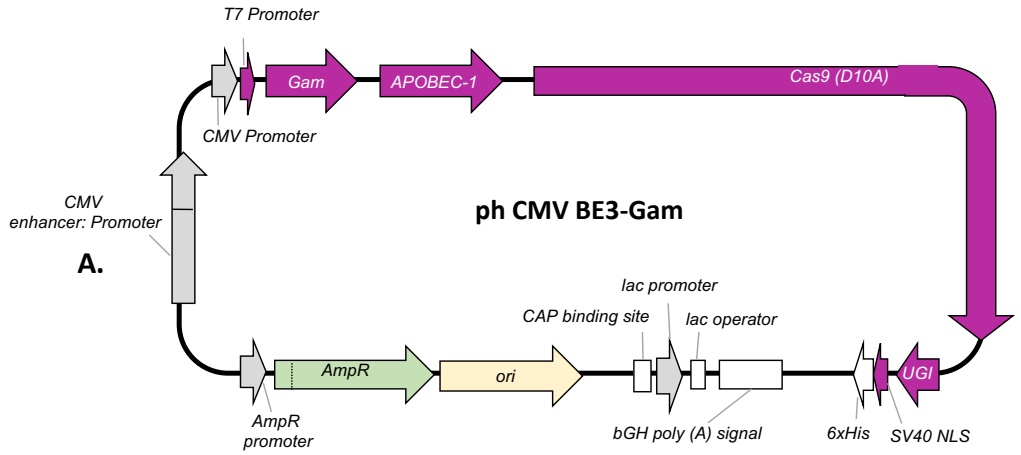
Human EPI-iPS cells were dissociated as previously described using TrypLE™ and $5x10^4$ cells were seeded into rhLaminin coated 12 wells with SF medium supplemented with Revitacell (1x) and cultured ON. The following day of transfection SF medium was aspirated and replaced with 500uL IMDM medium, 5µg plasmid (BE (3.75µg): u6gRNA(1.25µg)) were added to 2µL Lipofectamine Stem reagent (ThermoFisher) in IMDM and incubated at RT for 10'. The resulting complex was added to EPI-iPS cells and incubated at 37ºC for 4hrs, 500µL of prewarmed SF media was added to the culture and incubated at 37ºC ON. The following day all medium was aspirated from the wells and fresh SF media added.

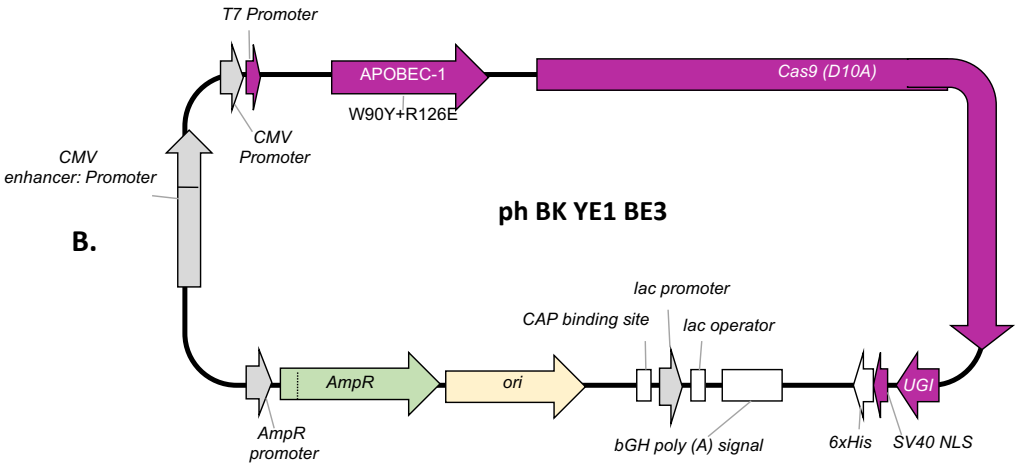
After 48hrs transfected cells containing the U6gRNA plasmid were selected with 450mg/ml G418 (Sigma). Five days later transfected cells were sorted into 96 well plates as previously described in **section 2.** After 7-14 days in culture gDNA was extracted from expanded clones for downstream analysis of editing efficiency.

2.9.4 TaqMan SNP Genotyping Assay

To assess single base editing in our transfected cells the TaqMan SNP genotyping assay was carried out (ThermoFisher). Two allele specific TaqMan probes were designed containing distinct florescent dyes and a PCR primer pair to detect specific GATA2 mutations. Annealed oligonucleotides for each mutation and wildtype allele sequence were used as reference samples along with a sample without gDNA as negative control (no template control (NTC)). Each reaction was set up in duplicate with 5ng gDNA in a 10mL final volume according to manufacturer's instructions. Reactions were carried out in a QuantStudio™ 7 Flex Real-Time

PCR System (Applied biosystems). After PCR amplification a post PCR plate read was carried out whereby the software uses the fluorescent measurements to plot R_n values in order to determine which alleles are in each sample. The Taqman Genotyper software was used to plot the allelic discrimination data as a scatter plot of allele 1 (Mutant) and allele 2 (wild type).





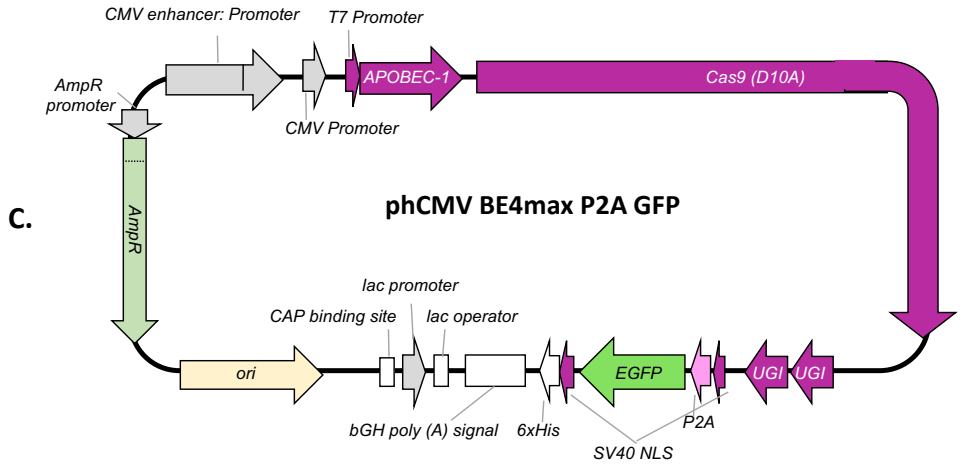


Figure 2.2 Vector map of plasmids used in gene editing experiments. CMV enhancer- enhancer of human cytomegalovirus human cytomegalovirus; CMV promoter- Strong mammalian expression promoter from the human cytomegalovirus human cytomegalovirus; T7 promoter-promoter for bacteriophage T7 RNA polymerase; APOBEC-1- cytidine deaminase enzyme derived from rat, facilitates C to U nucleotide editing; APOBEC-1 containing mutations in residues involved in APOBEC1 substrate binding (W90Y+R126E) that result in narrowed editing window width of approximately 2 nucleotides; Cas9(D10A)- nickase mutant of the Cas9 endonuclease from the *Streptococcus pyogenes* Type II CRISPR/Cas system, generates RNA-guided single strand nicks in DNA due to the D10A mutation in the RuvC catalytic domain; UGI- uracil-DNA glycosylase inhibitor from a Bacillus subtilis bacteriophage; SV40 NLS- nuclear localization signal of SV40 (simian virus 40) large T antigen, tags protein for import into cell nucleus by nuclear transport; P2A- 2A peptide from porcine teschovirus-1 polyprotein, allows translation of multiple genes in a single transcript; EGFP- enhanced green fluorescence protein, selectable marker to identify transfected cells; 6xHis- polyhistidine affinity tag composing six histidine residues; bGH poly(A) signal- bovine growth hormone polyadenylation signal, termination signal that initiates the process of releasing the newly synthesized RNA from the transcription machinery; AmpR promoter, drives transcription of the AmpR gene; AmpR- gene conferring resistance to ampicillin in bacteria, ori- origin of replication, DNA sequence which directs initiation of plasmid replication; CAP binding site- catabolite activator protein binding activates transcription in the presence of cAMP; lac promoter- promoter for the E. coli; lac operator- The lac repressor binds to the lac operator to inhibit transcription in E. col.

2.10 shRNA mediated KD in CB CD34^{+} cells

2.10.1 Viral Plasmids

psPAX2 and VSVG plasmids were kindly gifted to our lab by Professor Kamil Kranc (CRUK Barts Centre), psi-LVRU6GP plasmids containing shGATA2 were purchased from Genecopoeia. shGATA2 knockdown sequences are located in **Table 2.5** and plasmid details can be found in **Figure 2.3.** Plasmids were prepared from bacterial stabs as previously described in **section 2** using 100mg/mL ampicillin LB and agar plates.

Table 2.5: shRNA sequences GATA2, exon location following (ENSG00000179348)

shRNA	ID	Target Sequence	Location
shSCR GFP	CSHCTR001	5’ GCTTCGCGCCGTAGTCTTA 3’	N/A
GATA2	HSH007152-31	5’ GCAAATTGTCAGACGACAACC 3’	Exon 5 11272-11292
	HSH007152-32	5’ GGACAACTTTATGTAGAGAAA 3’	Exon 6 12417-12437
	HSH007152-34	5’ CCTTTAAAGTGAGTACTGTTA 3’	Exon 6 13714-13734

2.10.2 Generation of lentiviral particles

A second generation Lentiviral system was used in order to produce lentiviral particles using the Calcium phosphate method (Kingston, Chen and Rose, 2003). Lentiviral transfer vectors containing shRNA constructs against GATA2, ASXL1 or scramble control sequence were

diluted in dH_20 together with glycoprotein envelop plasmid VSV-G and packaging construct psPAX2 and mixed with calcium chloride $CaCl_2$ (2M). The solution was added dropwise to 2x HEPES buffered saline (HBS) and incubated at room temperature for 15'. The medium of HEK293T cells at 85-90% confluency was replenished containing 25mM Chloroquine and the incubated solution added dropwise to the cells. 6 hours later the medium was replenished and supernatant containing lentiviruses collected every 12 hours and passed through a 0.45mm filter.

Virus particles were concentrated at 23,000rpm at 4°C for 2.5hrs in a L8-70M ultracentrifuge, SW32Ti rotor (Beckman), resuspended in 50µL IMDM medium, snap frozen in dry ice and stored at -80°C.

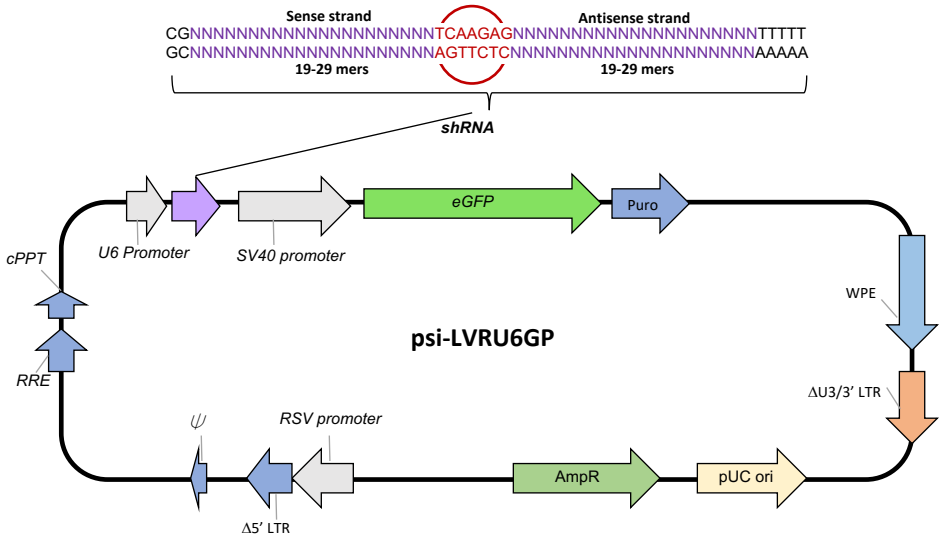


Figure 2.3: Vector map of psi-LVU6GP plasmids used in shRNA experiments. shRNA GATA2 sequences are located in **Table 2.7**. U6 is the human U6 small nuclear 1 promoter (polymerase III promoter); Simian virus 40 (SV40) allows the transcription termination and polyadenylation of mRNA transcribed by RNA polymerase II; enhanced green fluorescent protein (eGFP) allows for the detection and isolation of cells transduced with shRNA; Encephalomyocarditis virus internal ribosome entry site (IRES) recruits ribosome to initiate translation internally on a transcript independent of its 5' end. Multiple proteins can be made from a polycistronic transcript containing multiple ORFs separated by IRES; puromycin resistance gene (puro) allows cells to be resistant to puromycin; woodchuck hepatitis virus posttranscriptional regulatory element (WPE) enhances viral RNA stability in packaging cells leading to higher titre of packaged virus; truncated HIV-1 3' long terminal repeat (3'-LTR) allows packaging of viral RNA into virus and self-inactivate the 5'LTR by a copying mechanism during viral genome integration, contains polyadenylation signal for transcription termination; pUC origin of replication (pUC ori) facilitates plasmid replication in E. coli leading to high-copy plasmid number; Ampicillin resistance gene (Amp') allows E. coli to be resistant to ampicillin; truncated HIV-1 5' long terminal repeat (5' LTR) allows transcription of viral RNA and its packaging into virus; HIV packaging signal (Ψ) allows packaging of viral RNA into virus; HIV-1 Rev response element (RRE) allows rev-dependent nuclear export of viral RNA during viral packaging; central polypurine tract (cPPT) facilitates the nuclear import of HIV-1 cDNA through a central DNA flap; the Human phosphoglycerate kinase (hPGK) promoter allows high-level expression of the selection marker in mammalian cell lines and the mCherry and blasticidin dual reporter gene allows the selection of stable expression clones by mCherry fluorescent or blasticidin resistance.

2.10.3 Titration of lentiviral particles

To calculate virus titres HEK293T were seeded at 50,000 cells/ 24well and virus of different serial dilutions (1, 0.1, 0.01 and 0.001ml) added to the supernatant. After 24hrs HEK293T medium was replaced with fresh medium and 96 hours later cells were analysed by flow cytometry for transduction efficiency measured by selectable marker expression (eGFP). Titres were calculated from wells expressing between 10-30% positivity as higher transduction rates lead to multiple integration per cell and therefore an underestimation of the titre (Fehse *et al.*, 2004). To reach the number of transduced units (TU) the number of cells plated were multiplied by the percentage of positively transduced cells and divided by the volume of virus added to the supernatant (mL).

$$\textbf{TU/mL} = \frac{(5 \times 10^4) \times (\%\ \text{of eGFP/mCherry cells})}{(\text{volume of virus supernatant})}$$

The number of transduced units were then multiplied by the multiplicity of infection (MOI).

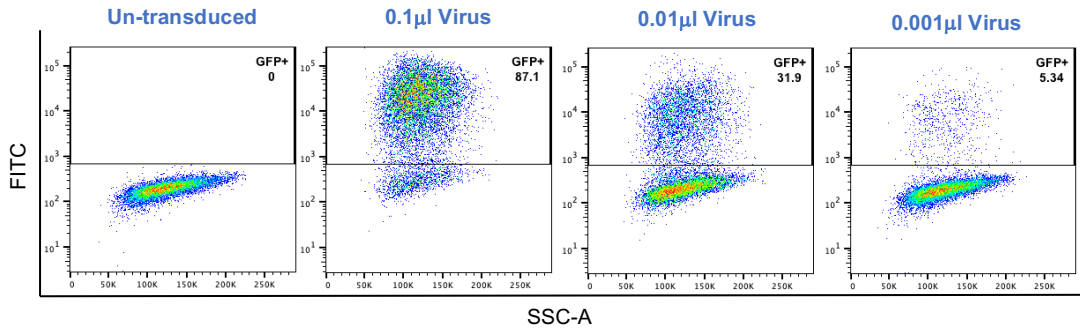


Figure 2.4 Gating strategy for lentivirus titration: 5x10^5 HEK29T cells were transduced with 0.1, 0.01 or 0.001mL of concentrated virus and GFP expression measured by flow cytometry against an untransduced control.

2.10.4 Primary Cord Blood cell isolation

Fresh cord blood samples were obtained from the Anthony Nolan Cell Therapy Centre (London). The blood was diluted 1:1 with Hanks Balanced Salt Solution (HBSS) supplemented with HEPES, Penicillin-Streptomycin (P/S) and bacteriocide-free heparin (1 unit/mL). To estimate the total number of CB-derived CD34$^+$ HSPC, 10µL of unprocessed blood was resuspended with 190ml RPMI 1640 and 1.5ml Zaponin and phenotype and purity assessed by two colour immunophenotypic analysis using anti-CD45 PerCP (Biolegend, UK) and CD34-PE (BD Biosciences, UK)

Platelets were depleted by centrifugation over Ficoll. Briefly, every 8mL of blood was layered on top of 5mL of Ficoll and centrifuged at room temperature at 400xg for 40'. Mononuclear cells (MNC)-visible as a clean interphase between the plasma and Ficoll, were collected,

washed with 15mL RPMI 1640 (supplemented with 5% FBS and bacteriocide-free heparin) and centrifuged at 1200rpm for 10'. The supernatant was removed, all pellets combined and washing steps repeated (20mL medium) until the supernatant was no longer hazy- indicating no platelet contamination. MNC cells were counted and resuspended in equal volume of cold 2x strength freezing medium (20%DMSO, 30% FCS, 50% IMDM medium) giving a final volume of 1mL/vial at approximately $5x10^7$ cells/mL.

2.10.5 CD34 manual enrichment

Highly enriched $CD34^+$ cells were derived from CB MNC using the Indirect CD34 Microbead kit MiniMACS (Miltenyi Biotec, Surrey, UK). Briefly, 1mL of FCS and 20mL DNase was added to each vial of frozen MNC and rapidly thawed at 37°C. The contents of the vial were aspirated into clean 50mL tube and an equal volume of room temperature PBS (supplemented with 5mM $MgCl_2$) was added dropwise so that the volume doubled over 3' until 3 doublings were reached (total volume 8mL). The cells were centrifuged at 1000rmp for 10', supernatant removed and 10mL DNase added to prevent cells from clumping. The pellet was resuspended in 400ml cold column buffer (1xPBS, 0.5% BSA, 5mM $MgCl_2$) /10^8cells based on cell count calculated post freezing. 100mL of FcR blocking reagent and 100ml of CD34-Hapten-Antibody per 10^8 cells were added and incubated at 4°C for 15'. 5mL of column buffer was added per 10^8 cells and centrifuged at 1200rpm for 5' (or 10' alternatively for volumes over 5mL). The supernatant was removed, pellet dispersed and resuspended in a final volume of 400ml/ 10^8 cells. 100mL of Anti-Hapten CD34 microbeads were added per 10^8 cells and incubated at 4°C for 15'. Next, 5mL column buffer was added to the cells, passed through a 70mL cell strainer, to remove any clumping cells, and centrifuged at 1200rpm for 5'. The supernatant was removed, cells resuspended in 500mL of column buffer and applied to a MS column complete with flow restrictor and held in place by a magnetic holder with a clean 15mL tube to collect the CD34- cells. Once column flow had stopped the column was washed four times with 500mL column buffer. The column was removed from the magnet and placed into a clean 15mL tube, 1mL column buffer was added to the column and the CD34+ cells eluted by pushing the plunger firmly into the top of the column. To ensure purity of >90% the eluted fraction was applied to a second column and enrichment repeated as previously described. Finally, 10mL of eluted cells was count, cells centrifuged at 1200rpm for 5' and subsequently cultured overnight at $2x10^5$ cells/mL in StemSpanII medium supplemented with hSCF (100ng/mL), hFLT3L (100ng/mL), hIL-6 (10ng/mL), hIL-3 (10ng/mL), G-CSF (10ng/mL), P/S (100U/mL).

2.10.6 Lentiviral transduction of CD34$^+$ HPSC

Human CD34$^+$ HSPC's were resuspended in stemness maintenance medium (StemSpanII medium supplemented with hSCF (300ng/mL), hFLT3L (300ng/mL), hTPO (20ng/mL) P/S (100U/mL) and HEPES (10mM) and transduced with lentiviral vectors encoding either shGATA2 (31, 32, 34 or scramble control) harbouring a eGFP selectable marker under the control of the SV40 promoter. 8 hours post transduction, CD34$^+$ HSPC's were collected, centrifuged at 12000rpm for 5', resuspended in fresh media and expanded for 4 days. On day 4 cells were stained with CD34-APC and live cells sorted based on CD34$^+$ and eGFP positivity.

2.10.7 Protein isolation and cell lysate preparation

Sorted eGFP or CD34$^+$ CB cells were pelleted by centrifugation at 400xg for 5', washed with cold PBS (1x) and re-centrifuged following the same conditions. Cell pellets (~1-2x 10^6 cells) were resuspended in 100mL lysis buffer (RIPA 1:100 dilution of protease/phosphatase Inhibitor Cocktail- Cell Signalling Technology). Samples were incubated on ice for 30' (vortexing at 10' intervals to facilitate cell lysis) and centrifuged at 4°C for 30' at 16,000xg. Proteins within the supernatant were collected, aliquoted (to prevent repeated freeze thaw cycles) and stored at -80°C.

2.10.8 Protein quantification

The total amount of protein per sample was quantified using the Pierceä BCA Protein Assay Kit (Thermoscientific) based on bicinchoninic acid (BCA) for the colorimetric detection and quantification of total protein. A set of protein standards were prepared by diluting one Albumin Standard- BSA (2mg/mL) with RIPA buffer according to **(Table 2.6)** 10mL of each standard or 10ml of diluted sample lysate (1:5 and 1:10 diluted in RIPA buffer) were added into a 96 well plate along with 200ml BSA working reagent (BCA reagent A and B prepared in a 50:1 ratio). The plate was covered from the light mixed thoroughly on a shaker for 30'' and incubated at 37°C for 30'. The absorbance of each sample/standard was read at 562nm using a CLARIOstar microplate reader (BMG Labtech. Finally, the absorbance values of protein standards were used to create a standard curve to calculate the protein concentration of each unknown sample.

Table 2.6: Pierce™ BCA Protein Assay standard and sample preparation.

Standard	Volume of RIPA (μL)	Volume and Stock of BSA (μL)	Final BSA concentration (μg/mL)	Volume (μL) /96 well
A	0	300 Stock	2000	10
B	125	375 Stock	1500	10
C	325	325 Stock	1000	10
D	175	175 vial B dilution	750	10
E	325	325 vial C dilution	500	10
F	325	325 vial E dilution	250	10
G	325	325 vial F dilution	125	10
H	400	100 vial G dilution	25	10
I	400	0	0= Blank	10

	Dilution	Volume (μL)/96 well
Sample	1:5/1:10	10

2.10.9 Protein separation and transfer to blot

10μg of protein lysate was mixed with to a final 1x concentration with 4x Laemmli buffer (200mM Tris pH 6.8, 10% SDS, 40% glycerol, 8% β-Mecaptoethanol, 0.05% Bromophenol blue) and denatured at 95°C for 5'. A 4-15% Mini-PROTEAN TGX precast gel (BioRad) was assembled into an electrode chamber and placed into an electrophoresis tank and both the chamber and tank filled with running buffer (25mM Tris, 190mM Glycine, 0.1% SDS). 7μL of protein ladder was added to the first well followed by 20μL of each denatured sample. The gel was run at 120V for 45' to 1hr. Proteins were transferred using a Trans-Blot Turbo Mini PVDF transfer pack consisting of two buffer saturated ion reservoir stacks and a prewetted 0.2 μm PVDF membrane, using Trans-Blot Turbo Transfer System. Briefly, the bottom iron reservoir stack is placed in contact with the bottom cassette electrode (anode) with the blotting PVDF membrane on top. The gel is placed face down, with the top ion reservoir stack in contact with the top cassette electrode (cathode). The complete cassette was locked, placed into the Trans-Blot Turbo and proteins transferred to the membrane for 7' at 2.5A and 20V.

2.10.10 Protein blotting and visualisation

Following protein transfer the PVDF membrane was blocked using blocking buffer (TBS/T (37mM NaCl, 20Mm Tris pH 7.6, 0.3% Tween) with 5% BSA) for 1hr at room temperature. The membrane was probed with primary antibody specific to the protein of interest **(Table 2.7.)** in blocking solution at 4°C overnight. Membranes were washed with TBS/T initially for 15', following by two 5' washes and incubated with secondary antibody **(Table 2.6)** for 1hr at room temperature. Washing steps were repeat as described above and membrane incubated in the dark with Immobilon Forte Western Horseradish peroxide (HRP) substrate (Merck) for 5-10' (depending on antibody). Residual substrate was removed from the membrane and the chemiluminescent signal detected using the ChemiDoc MP Imaging system (BioRad) whereby

proteins are visualised by the chemiluminescence produced by the chemical substrate being catalysed by the HRP conjugated to the secondary antibody. Protein expression was normalised to housekeeping gene GAPDH and Image J software used to quantify differences in band intensity between the protein of interest and the housekeeping gene.

Table 2.7: Primary and secondary antibodies used during western blotting protocol.

Primary Antibodies				
Antigen	**Host**	**Clone**	**Dilution**	**Supplier**
GATA2	Rabbit IgG		1:2000	Abcam
GAPDH	Mouse IgG1	6C5	1:10000	Mereck MAB374

Secondary Antibodies			
Antigen	**Host**	**dilution**	**Supplier**
Anti-Mouse IgG1 (HRP)	Goat	1:5000	Abcam 98693
Anti-rabbit IgG (HRP)	Goat	1:5000	Abcam 205718

2.10.11 Human CFU Assay

600 sorted $CD34^+$ CB cells expressing eGFP were plated in duplicate into 1mL of methycellulose H4434 supplemented with 100U/mL P/S (Stemcell Technology) and 10µg/mL lipopolysaccharide (LPS) (*Escherichia coli* O55:B5) (Sigma) into 6wells. Colonies were scored and harvested on day 10 of culture. Cultures were collected with IMDM (2% FBS) using a 5mL syringe, centrifuged at 300xg for 5' at RT resuspended in 1mL of IMDM (2% FBS) for counting. Cell concentration was readjusted to $2.5x10^5$ cells and plated into secondary CFC assays as outlined above treated with 10µg/mL LPS. Secondary CFCs were scored at day 10.

2.10.12 $CD34^+$ CB myeloid expansion with LPS challenge

Sorted $CD34^+$ HSPC's were maintained in myeloid expansion medium (StemSpanII medium supplemented with hSCF (50ng/mL), hFlt3L (50ng/ml), hIL-3 (50ng/mL), hIL-6 (25ng/mL), hG-CSF (25ng/mL), hGM-CSF (25ng/mL) andP/S (100U/mL)), treated with (LPS) for two weeks to promote the expansion and differentiation of human myeloid progenitor cells. Cells were harvested at day 7 and 14 for downstream analysis.

2.10.13 Detection of apoptosis by annexin V

Cells were incubated in annexin binding buffer (BD biosciences) containing annexin V antibody (BioLegend) and incubated for 25' in the dark. To stop the reaction annexin binding buffer was added and 2µL of DAPI (20µg/mL) added prior analysis. Annexin V is a 35-36 kDa Ca2+-dependent phospholipid binding protein with a high affinity for anionic phospholipid phosphatidylserine (PS). Within healthy cells PS is located on the cytoplasmic surface of the plasma membrane, during apoptosis the plasma membrane undergoes conformational changes resulting in translocation of PS to the extracellular side of the plasma membrane. At

late stages of apoptosis membrane integrity is compromised allowing DAPI to enter the cell and stain the nuclei (Demchenko, 2013) **(Figure 2.5).**

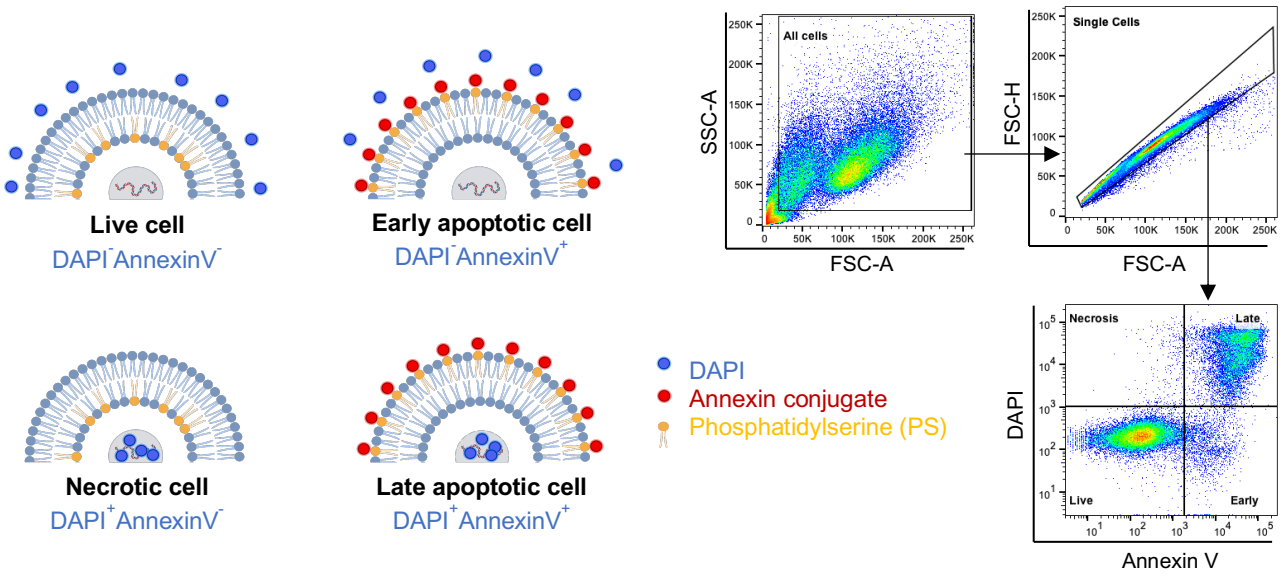


Figure 2.5: Annexin V detection of apoptosis. Schematic of molecular events during apoptosis pathway and representative FACS plots showing: Live cells ($DAPI^-AnnexinV^-$) Early apoptotic cell ($DAPI^-AnnexinV^+$), Late apoptotic cell $DAPI^+AnnexinV^+$) and Necrotic cell ($DAPI^+AnnexinV^-$).

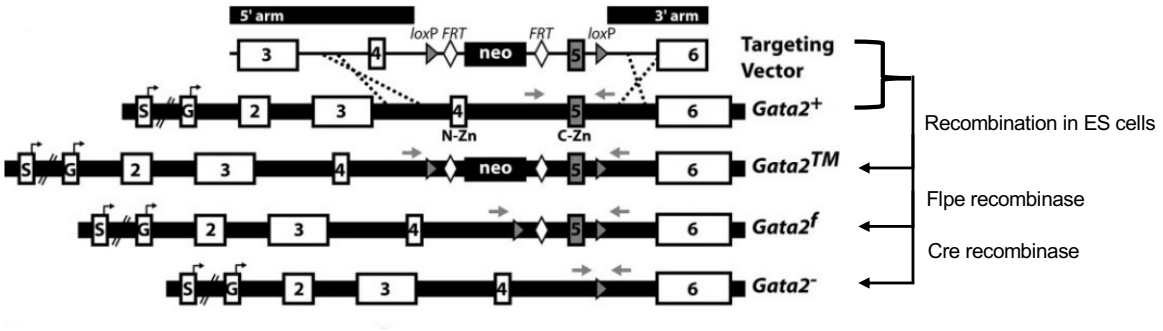
2.11 Mice

All experimental mice used were sex-matched 6–8-week-old mice of the congenic C57BL/6 strain. All animal experiments were authorised by the UK Home Office under project licence (30/3380).

2.11.1 *Gata2* mouse strain (*Gata2*^{tm1Sac})

The *Gata2* transgenic mouse was kindly gifted to our lab by Professor Julian Downward (The Francis Crick Institute, London). These animals were engineered in the laboratory of Sally Camper, where the *Gata2* locus was modified to create an allele that could be utilised for tissue-specific and inducible gene deletion (Charles *et al.*, 2006). A targeting vector containing a neomycin resistant cassette flanked by *FRT* sites located in the intron upstream of exon five-encoding the C-terminal ZF domain- and *loxP* sites at the 5' of the neomycin cassette and a second at the 3' of the fifth exon were inserted into the *Gata2* locus in ES cells which upon homologous recombination resulted in the production of the *Gata2*^{tm1Sac} allele (Gata2^{TM}). These ES cells were used to produce germline mouse chimeras of which *Gata2*^{TM} heterozygotes were mated to Flpe recombinase transgenic mice resulting in the removal of the neomycin cassette, creating the floxed allele - *Gata2*^{f}. Mating of mice with the *Gata2*^{f} allele

to Cre transgenic mice (in specific lineages or in an inducible manner) results in the excision of the fifth exon giving rise to a non-functional truncated *Gata2* protein (Charles *et al.*, 2006) (**Figure 2.6**).



Modified from Charles et al. 2006

Figure 2.6: Strategy used for the engineering of the Gata2^{tm1sac} transgenic mouse model. The fifth exon contains the C-terminal zinc finger domain; thus deletion of this exon abolishes the DNA binding ability of Gata2. A targeting vector containing a neomycin resistant cassette flanked by FRT sites and loxP sites flanking the fifth exon were inserted into the Gata2 locus in ES cells. Upon Flpe recombinase activity the neomycin cassette was excised generating the Gata2^{fl} allele.

2.11.2 *Vav-iCre* breeding with *Gata2 ^{tm1Sac}* and induction of Gata2 deletion *in vivo.*

The *Vav-iCre* transgenic mouse was kindly gifted to our laboratory by Professor Kamil Kranc (Queen Mary, University of London) and has been described previously**.** These transgenic mice express Cre recombinase under the control of the mouse vav1 oncogene (Vav1) promoter. The *Vav-iCre* or improved Cre, is an optimisation of the Cre gene for optimised codon usage, removed putative cryptic splice sites and reduced CpG content to limit epigenetic silencing (Shimshek *et al.*, 2002). Both the *Vav and Vav-iCre* are driven from the *murine Vav1* promoter and disrupt *floxed* genes in fetal and adult HSCs and their descendants, along with reports of *Vav-iCre* transgene expression occurring in the testis(de Boer *et al.*, 2003). To avoid or minimize germline deletion of the floxed allele it is essential to breed *Vav-iCre* females with floxed males. *Vav-iCre$^{-/-}$;Gata2$^{f/f}$* were bred with *Vav-iCre$^{+/+}$;Gata2$^{+/+}$* to obtain *Vav-iCre$^{+/-}$;Gata2$^{fl/fl}$* and control *Vav-iCre$^{-/-}$;Gata2$^{fl/fl}$* mice.

2.11.3 DNA extraction and genotyping of animals

DNA was extracted from ear punch samples using the ISOLATE II Genomic DNA kit. Ear samples were incubated in lysis buffer GL and proteinase K for 2-3 hours at 56ºC, lysis buffer G3 was subsequently added to achieve the final lysate and further incubated for 10 minutes

at 70ºC. DNA extraction was continued as described in **section 2.9.3** from the addition of 100% ethanol to adjust the binding conditions.

Genomic PCR was carried out following manufacturers protocol (MangoMix™, Bioline). DNA primers and PCR settings are found in **Table 2.8**. Electrophoresis was carried out using a 2% [w/v] 1xTris-acetate-EDTA (TAE) agarose gel ran at 120V for 40'.

Table 2.8: Genotyping primers and cycling conditions.

Primer	Sequence	PCR programme
Gata2 Fwd	5'-GCCTGCGTCCTCCAACACCTCTAA-3'	1. 94ºC 5 minutes 2. 94ºC 1 minute 3. 60ºC 1 minute 4. 72ºC 1 minute Repeat steps 2-4 for 34 cycles 5. 72ºC 10 minutes 6. 4ºC hold
Gata2 Rev	5'-TCCGTGGGACCTGTTTCCTTAC-3'	
Vav-iCre Fwd	5' CCGAGGGGCCAAGTGAGAGG 3'	1. 94°C,40 sec 2. 64°C, 40 sec 3. 72°C, 30 sec 4. 72°C, 5 min Repeat steps 2-4 for 30 cycles 5. 72°C, 5 min 6. 4°C hold
Vav-iCre Rev	5' GGAGGGCAGGCAGGTTTTGGTC 3'	

2.11.4 Induction of Polyinosinic:polycytidylic acid (poly I:C)

Inflammation was induced in 6–8-week-old gender-matched mice by inteperitoneal (I.P) administration of 10µg/g body mass of poly I:C (GE Healthcare) in PBS at two-day intervals for up to seven days.

2.11.5 Dissection and PB extraction

On the seventh day peripheral blood (PB) was obtained by nicking the tail vein with a scalpel followed by collection of 30-50µL of blood in EDTA-coated bleeding tubes (Starstedt). Subsequently, mice were culled by cervical dislocation and the spleen, two femur and two tibiae aseptically removed and placed in DPBS (Gibco) with 2% FBS.

2.11.6 Red blood lysis

Ammonium chloride (NH_4Cl) (StemCell technologies) was used to lyse red blood cells from the PB. During hypotonic conditions (0.8% NH_4Cl) H_20 enters RBCs as they have weaker membrane tonicity in comparison to nucleated RBCs or leukocytes, resulting in osmotic lysis

of the cells. 12µL of blood was added to 600µL of NH_4Cl, incubated for 6', mixed by repetitive inversion of the tube and further incubated for an additional 6'. Samples were centrifuged at 370 xg for 10' at RT, followed by removal of the supernatant containing lysed RBCs leaving a WBC pellet.

2.11.7 Bone marrow extraction

After cleaning, bones were crushed in DPBS with 2% FBS using a pestle and mortar. Cells were collected in DPBS with 2% FBS and filtered using a 70µm strainer (Miltenyi Biotec) to remove cell clumps. This step was repeated twice by washing the crushed bones with DPBS with 2% FBS to obtain a 30mL single cell suspension of total BM cells.

2.11.8 Splenocyte extraction

Spleens were placed in a 70µm strainer, inside a 6cm petri dish with DPBS with 2% FBS. The spleen was homogenised using a 5mL syringe plunger, cell suspension resuspended by pipette to remove clumps and repeated twice to achieve a 7mL spleen single cell suspension.

2.11.9 Murine Colony forming cell (CFC) assay

CFC assays were carried out according to the manufacturers protocol. Briefly, 1-2x10^4 total BM cells were resuspended in 250uL of IMDM media, added to Methocult™ M3434 semi solid media (Stemcell technologies), vortexed briefly and plated in duplicate 6 wells using a 2 mL syringe and 19G blunt end needle. Colonies were identified and counted using an inverted microscope (Leica) routinely at day 10.

2.12 Flow Cytometry

2.12.1 Intracellular staining

Cells at 70% confluency were washed with DPBS, dissociated with trypsin and centrifuged to form a single cell suspension. In order to exclude dead cells during intracellular analysis the Zombie viability dye (ThermoFisher) was used prior to fixation. Zombie is an amine-reactive fluorescent dye that is permanent to cells with compromised membranes and non-permanent to live cells. Cells were resuspended in a 1:100 dilution of Zombie dye to DPBS per 1x10^6 cells and incubated at RT for 15 mins in the dark. The suspension was diluted in DPBS 2% FBS, centrifuged for 5 mins at 300xg, washed with DPBS to remove traces of serum. Cells were fixed in 0.5% methanol free Paraformaldehyde (PFA) (ThermoFisher) for 10' at room temperature and permeabilized using 0.1% Tritonx100 (Sigma), 2% BSA in PBS for 15'. Samples were incubated with fluorescence-labelled primary antibodies for 30' in the dark, washed twice and resuspended according to pellet size in 2% FBS in PBS. Samples incubated

with purified primary antibodies were washed and subsequently incubated with fluorescence-labelled secondary antibodies for 30' in the dark. Processed cells were finally filtered through a 30μM strainer and analysed.

2.12.2 Extracellular staining

Cells were harvested to a single cell suspension, pelleted by centrifugation and washed with DPBS. Extracellular staining was carried out in DPBS 2% FBS with fluorescence labelled primary antibodies at 4°C in the dark. Cells were washed with DPBS 2% FBS, filtered through a 30μM strainer and 20μg/mL DAPI added to suspension prior to analysis.

2.13 Cell sorting

2.13.1 Haematopoietic Progenitor Cell (HPC) sorting

Enriched $CD34^+$ populations from OP9/PSC cultures or cells differentiated using the StemDiff™ Haematopoietic Kit were stained with antibodies to detect Haematopoietic progenitor cells (HPCs) according to previous publications, immunophenotypically characterised as $DAPI^-CD34^+CD31^+CD43^+$ (Chadwick *et al.*, 2003; Choi, Vodyanik and Slukvin, 2011) and sorted using an 85μm nozzle.

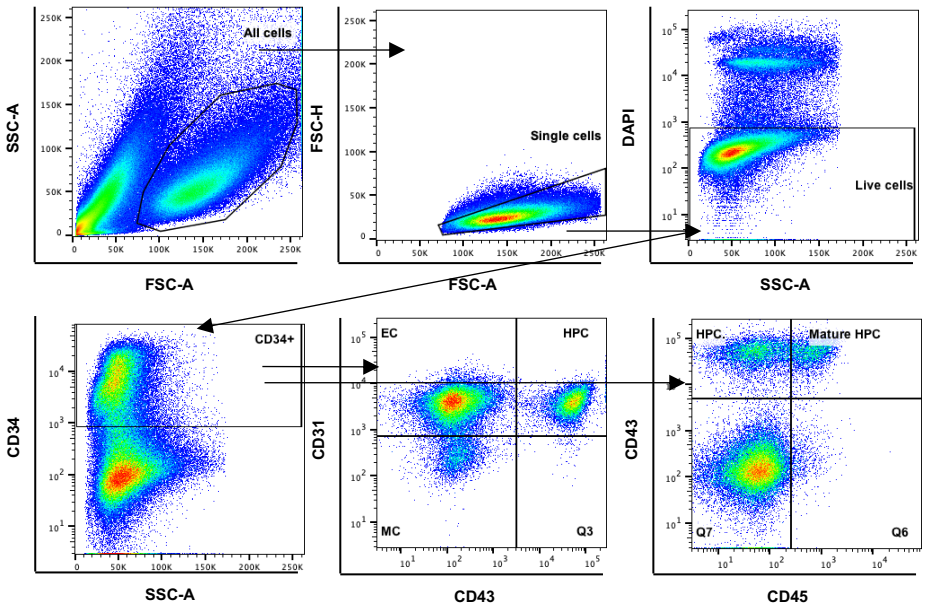


Figure 2.7: Gating strategy used for Haematopoietic differentiation of pluripotent stem (PSC) cells using the StemDiff™ culture system.

2.13.2 Single cell sorting into 96 wells

Seventy-two hours post electroporation; cells were dissociated with TrypLE™, washed twice with DPBS and stained with TRA-1-60-R for 20' at 4°C in the dark. Prior to sorting, samples were filtered through a 30-µm filter to remove cell clumps and 20µg/mL DAPI added for live dead staining. Single viable cells were sorted into each well of a rhLaminin coated (0.5-2.0 µg/cm^2) 96 well plates, containing complete SF medium (supplemented with Penicillin-Steptomycin (100U/mL) and 1x RevitaCell™) using a 100µm nozzle, under sterile conditions **(Figure 2.8).** After sorting, plates were centrifuged at 70xg for 3' at room temperature to facilitate seeding and incubated at 37°C in 5% CO_2.

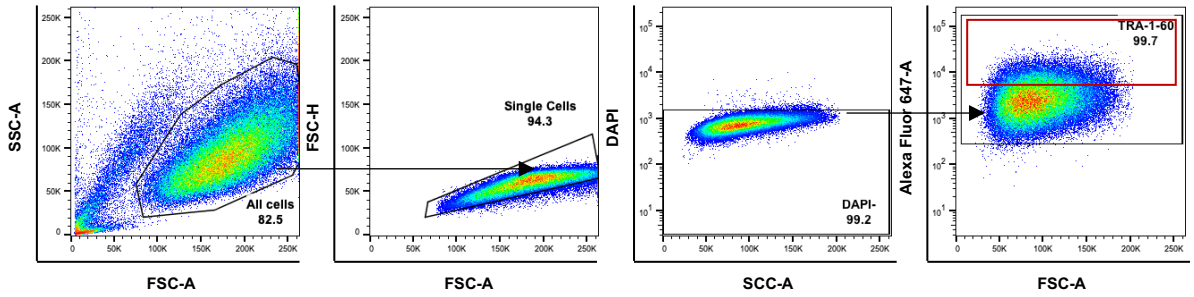


Figure 2.8: Gating strategy for single cell sorting of TRA-1-60 positive human iPS cells. TRA-1-60 positive human iPS cells were transfected by electroporation with TrueGuide™Synthetic gRNA containing T354M sequence and CAS9 Ribonucleoproteins (RNP). Cells were gated from live cells (DAPI-) and the top 50% TRA-1-60$^+$ cells were sorted into each 96 well.

2.14 Flow cytometry Analysis

Flow cytometry was performed on a 4 laser BD LSRFortessa™ analyser and all sorting experiments conducted using the BD FACS Aria Fusion™ flow cytometer. See **Table 2.9 and Table 2.10** for full antibody information. Manual compensation was conducted in FACS Diva software using single stains for each anti-body associated fluorochrome using in each experiment. Where cell numbers were limited, single stains were carried out using UltraComp eBeads™ (ThermoFisher). Compensation was adjusted by plotting each fluorochrome against all other individual fluorochromes in the experimental panel and equalising the median fluorescence intensity (MFI) of the positive and negative populations **.** During PSC/OP9 culture analysis, fluorescence minus one (FMO) control were used in order to help set gating for each population. Gating strategies for murine BM and spleen analysis are outlines in **Chapter 5 section 5.3**.

2.15 Data and statistical Analysis

Data was subsequently analysed using FlowJo 10.6.1 software and results graphed using GraphPad Prism 8.4.1 (460). Data is presented as mean ± standard error of the mean (SEM).

Table 2.9: Mouse antibodies used for flow cytometry (FC).

Antigen	Clone	Reactive specie	Fluorochome	Concentration	Company
CD4	GK1.5	Mouse	Biotin, PE, PE-Cy7	1/1000	BioLegend
CD8a	53-6.7	Mouse	Biotin, PE, APC-Cy7	1/1000	BioLegend
CD9	MZ3	Mouse	PE	1/1000	BioLegend
CD11b	M1/70	Mouse/Human	Biotin, PE, APC	1/1000	BioLegend
CD14	M5E2	Human	Pacific blue	1/500	BioLegend
CD16/32	93	Mouse	PE-Cy7	1/25	BioLegend
CD19	6D5	Mouse	APC-Cy7	1/100	BioLegend
CD33	WM53	Human	APC	1/500	BD Biosciences
CD34	RAM34	Mouse	FITC, APC	1/25	eBiosciences
CD41	MWreg30	Mouse	FITC	1/1000	BioLegend
CD45.1	A20	Mouse	BV510, APC	1/500	BioLegend
CD45.2	104	Mouse	Pacific blue, BV510, PE	1/500	BioLegend
CD45R/B220	RA3-6B2	Mouse/Human	Biotin, FITC, APC	1/1000	BioLegend
CD48	HM48-1	Mouse	FITC	1/50	BioLegend
CD51	RMV-7	Mouse	PE	1/1000	BioLegend
CD71	RI7217	Mouse	PE	1/1000	BioLegend
CD105/Endoglin	MJ7/18	Mouse	PE	1/50	BioLegend
CD115/CSF-1R	AFS98	Mouse	APC-Cy7	1/50	BioLegend
CD117	2B8	Mouse	PE, APC	1/100	BioLegend
CD127/IL-7	A7R34	Mouse	BV650	1/50	BioLegend
CD135/Flt3	A2F10	Mouse	PE	1/50	BioLegend
CD150	TC15-12F12.2	Mouse	PE-Cy7	1/100	BioLegend
Gr1/Ly6-c	RB6-8C5	Mouse	Biotin, FITC, PE-Cy7	1/1000	BioLegend
Ly6-G	1A8	Mouse	PE-Cy7	1:1000	BioLegend
Nk1.1	PK136	Mouse	BV650	1/1000	BioLegend
Ter119	Ter119	Mouse	Biotin, APC-Cy7	1/1000	BioLegend
Sca-1	D7	Mouse	PE, APC-Cy7	1/25	BioLegend
Streptavidin		Mouse/Human	Pacific blue, PerCP	1/100	BioLegend, eBiosciences
γ-H2AX	2F3	Mouse/Human	APC/PE	1/200	BioLegend
Mouse IgG1, κ Isotype Ctrl	MOPC-21		APC	1/200	BioLegend
Fc block	93	Mouse		1/100	BioLegend

Table 2.10: Human antibodies used for flow-cytometry (FC) and immunofluorescence (IF).

Assay	Antigen	Clone	Fluorochrome	Concentration	Company
Pluripotency (IF)	OCT3/4	40/Oct-3	Purified	1/200	BD
	SOX2	O30-678	Purified		BD
	SSEA-1	HI98	Purified		BD
Pluripotency (FC)	OCT4	3A2A20	PE	1/1000	Biolegend
	SOX2	14A6A34	Alexa-Fluor-488		Biolegend
	TRA-1-60	TRA-1-60-R	Alexa-Fluor-647		Biolegend
	Isotype Ctrl IgG1k		Alexa-Fluor-488		BD
Hematopoietic differentiation (FC)	CD14	M5E2	PE		Biolegend
	CD15	W6D3	FITC		Biolegend
	CD235a	Hi264	PE-CY7		Biolegend
	CD29		PE-CY7		Biolegend
	CD31	WM59	FITC		BD
	CD34		APC		Biolegend
	CD43	1G10	PE		BD
	CD45	HI20	APC-CY7		Biolegend
	TLR-4	HTA125	APC		Biolegend
DNA damage (FC)	γH2AX	2F3	PE		Biolegend
	Isotype ctrl IgG1κ	MOPC-21	PE		Biolegend

Chapter 3

Characterisation of human iPS cell lines

3.1 Introduction

3.1.1 iPSC technology

Somatic cell reprogramming to induced pluripotent stem cells (iPSCs) was first realised by Dr Shinya Yamanaka, where he presented that transient exogenous expression of a combination of specific transcription factors; Oct4 (Pou5f1), Sox2, Klf4 and c-Myc (OSKM) could revert a mature or specialised cell back into an immature stem cell capable of self-renewal and differentiation into multiple cell lineages (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007). During the last decade many groups have sought to decipher the exact mechanism in which the reprogramming process occurs. Researchers have determined that mouse somatic reprogramming occurs as four main events; activation of alkaline phosphatase, silencing of somatic-specific expression, expression of SSEA1 and progressive silencing of exogenous genes in parallel to upregulation of endogenous Oct4 and Nanog. However, the understanding of human somatic reprogramming remains to be fully defined (Brambrink *et al.*, 2008; Stadtfeld *et al.*, 2008; David and Polo, 2014; Teshigawara *et al.*, 2017a).

3.1.1.1 Molecular mechanism of reprogramming

One of the initial parts of the reprogramming process is establishing the re-expression of genes that have been developmentally silenced which are located in condensed chromatin regions. Various studies have identified a specific class of TFs termed ‘pioneer’ TFs, that are able to access tightly packed chromatin structures where they are capable of inducing chromatin remodelling events, directing additional TFs and epigenetic remodelling enzymes to their location, whilst also being able to prevent methylation by protecting cysteine residues (Xu *et al.*, 2007; Zaret and Carroll, 2011; Hsu *et al.*, 2015). These events result in the progressive relaxing of chromatin during cell division, which exposes specific gene promoters that other TFs can directly access and bind (Krause, Sancho-Martinez and Izpisua Belmonte, 2016). Another key stage of murine fibroblast reprogramming is the mesenchymal-to-epithelial transition (MET), where epithelial genes such as *E-Cadherin, Cdh1, Epcam and epithelial associated mmu*-miRNA-200 family are upregulated though exogenous expression of Oct4 and Sox2 binding to promoter regions of *mmu-miRNA-200 family* (Wang *et al.*, 2013), whereas mesenchymal genes; *Snail1/2 Zeb1/2 and N-Cadherin* are downregulated (Krause, Sancho-Martinez and Izpisua Belmonte, 2016). This process has been defined as the mmu-miRNA-200/Zeb2 pathway which induces expression changes in thousands of genes (Esteban *et al.*, 2012), and is further supported by studies showing repression of mmu-miRNA using specific inhibitors results in repression of MET and iPS generation (Wang *et al.*, 2013). Moreover, it has been postulated that the intersection between EMT (a major driving force for

embryonic development) and MET is a central crossroad for cell fate decisions (Teshigawara *et al.*, 2017b).

Distinct differences in states of murine and human iPSCs have been observed, with murine iPSCs being in a 'naïve' state similar to murine ESCs. However, human iPSCs have been suggested to be in a 'primed' state, reflecting human ESC and murine Epistem cells (EpiSCs) that are more advanced pregastrulating stage embryo cells (Chia *et al.*, 2010; Hanna *et al.*, 2010; Jouneau *et al.*, 2012). The 'naïve' pluripotency state has also been shown to result in the reactivation of the X chromosome during female somatic cell reprogramming and is seen as an indication of pluripotency in its truest form. As human iPS cells do not represent this 'naïve' state this suggests that the reprogramming of murine and human somatic cells using the same 'OSKM' factors are unique from one another (Pasque and Plath, 2015). The conversion of a 'primed' state to a 'naïve' state in mice is facilitated by enforced expression of exogenous *Klf4* in specific ground state culture conditions (Guo *et al.*, 2009). In humans it has been proposed that this transition between states is regulated by pluripotency-associated has (Homo sapiens)- miRNA-290/303 family, however, the exact mechanism is still unknown (Gafni *et al.*, 2013; Gu *et al.*, 2016). In human ESCs *OCT4*, *NANOG* and *SOX2* co-occupy a large proportion of over 300 target genes which form regulatory circuits including autoregulatory and feedforward loops, suggesting that efficient transcription of these factors are essential for the stabilisation of a pluripotency network, which is further supported by their use as key factors in the reprogramming process (Boyer *et al.*, 2010).

3.1.1.2 Reprogramming factors and delivery

iPSC generation was initially achieved through the delivery of reprogramming factors delivered using retroviral/lentiviral expression vectors to achieve efficient reprogramming in a short timeframe. Although expression of these factors are silenced after the reprogramming of somatic cells to iPS cells occurs, the potential of reactivation poses potential complications for their clinical use in patient-specific therapy (Yao *et al.*, 2004; Hotta and Ellis, 2008). Many of the original reprogramming factors such as *CMYC* and *KLF4* have been shown to have oncogenic potential when overexpressed within a cell (Okita, Ichisaka and Yamanaka, 2007). Various studies have now described alternative reprogramming factors that can replace the original 'OSKM' cocktail described by Yamanaka. It seems that key up or downstream components of pathways regulated by the key pluripotent machinery or closely related protein family members, allow for the replacement of the 'Yamanaka factors'- highlighting the redundancy among their activities and target genes (Krause, Sancho-Martinez and Izpisua Belmonte, 2016). Furthermore, to overcome the dangers of random integration, technical advances have led to the development of several techniques that enable reprogramming

whilst minimizing DNA alterations **(Table 3.1).** Nevertheless, even with these advancements in methodological and technical improvements in iPSC generation over the past decade, reprogramming efficiency remains low- ~0.1% in humans and ~1.0% in mice (Stadtfeld and Hochedlinger, 2010).

Table 3.1: Delivery methods of reprogramming factors for induced pluripotent stem (iPS) cell generation

Integrated (with/without excision)	None-integrating
Retro/lenti viral vectors Multiple single gene/ polycistronic Single expression cassette	**Adenoviral vectors** Multiple single gene/ polycistronic
lenti viral vectors with Cre-mediated excision	**Sendai viral vectors**
lenti viral vectors with piggyBac (PB) transposon excision	**(Non-replicating)** Plasmids Mini circles
	(Replicating) oriP/Epstein-Barr nuclear antigen-1 (EBNA1) based episomes
	DNA-free RNA delivery Protein delivery

3.1.2 PSC haematopoietic differentiation

During embryogenesis distinct signalling pathways are activated or repressed in a sequential manner in order for efficient haematopoietic development to occur. During human pluripotent stem cell (PSC) differentiation, these intrinsic and extrinsic signal cascades are mimicked in order to generate cells throughout this process (Ackermann *et al.*, 2015). In the early phase of *in vitro* haematopoietic differentiation, PSC give rise to a primitive mesoderm (PM) population which express typical markers apelin receptor (APLNR) and platelet-derived growth factor receptor (PDGFR)α/CD140a. These cells upregulate meso/endodermal TFs *Brachyury (T), MIXL1* and *FOXF1* and have the potential to form hemangioblast (HB) colonies. These are specifically detected in serum free media supplemented with FGF2 and mirror their *in vivo* counterparts located in the posterior region of the primitive streak that express KDR and T (Slukvin, 2013a, 2016). Further differentiation of the PM population to hematovascular mesodermal progenitors (HVMP) ($KDR^{bright}APLNR^{+}PDGFR\alpha^{low/-}$) occurs through the downregulation of PDGFRα, along with downregulation of primitive streak genes and up-regulation of genes associated with angio-haematopoietic development (*TAL1, HHEX, LMO2,*

GATA2 and *ETV2*). This stage of differentiation marks the progressive commitment towards an endothelial and haematopoietic fate **(Figure 3.1 A)**.

The first endothelial progenitors that emerge from the HVMP, are characterised by expression of VE-cadherin/CD144 and CD31. This heterogeneous population can be subdivided into three populations with different potential; non-hemogenic endothelial progenitors (non-HEP), hemogenic endothelial progenitors (HEP) and angiogenic haematopoietic progenitors (AHP) based on CD73, CD43 and CD235a surface expression (Choi *et al.*, 2012) **(Figure 3.1 B)**. It has also been demonstrated that expression levels of *Runx1* can be used to define these populations, supported by reports that observe Runx1 expression to be critical for EHT during embryogenesis (Hirai *et al.*, 2003; Chen *et al.*, 2009). The earliest VE-cadherin^{+} population is the AHP which possess primary haematopoietic cell characteristics, though retains endothelial potential. These cells express a CD235a^{+}CD43^{low} phenotype and hold FGF2 haematopoietic cytokine-dependent erythromyeloid potential. The HEP is distinct from the non-HEP through the lack of CD73 and CD43 surface expression and high levels of GFI1 and RUNX1 (Choi *et al.*, 2012). These cells also have primary endothelial characteristics, lack haematopoietic CFC potential, though are able to generate blood cells upon co-culture with OP9 stromal layers. Non-HEPs have the capacity to segregate into endothelial cells with arterial or venous properties based on their expression of CXCR4 (Ditadi *et al.*, 2015). Multipotent haematopoietic progenitors (MHP) are derived from the HEP population and are associated with high expression of CD43 surface expression which give rise to distinct subsets of CD43^{+} haematopoietic cells such as erythro-megakaryocytic progenitors (CD41a^{+}CD235a^{+}) and multipotent myelolymphoid progenitors (lin^{-}CD34^{+}CD43^{+}CD45$^{+/-}$) (Vodyanik, Thomson and Slukvin, 2006; Choi *et al.*, 2009) **(Figure 3.1 C)**.

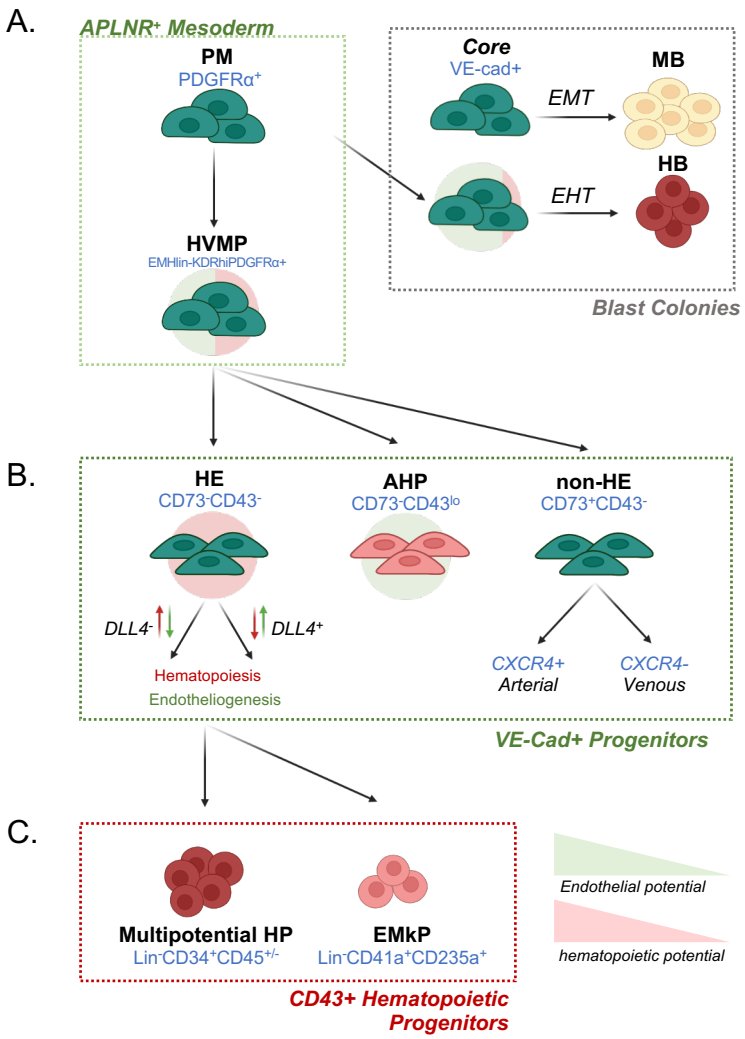


Figure 3.1 Stages of haematopoietic development from human PSCs. (A)hPSCs undergo mesoderm commitment marked by expression of mesodermal markers APLNR and KDR and lack of endothelial, endothelial/mesenchymal and haematopoietic markers (EMH lin- phenotype). The most primitive APLNR$^+$PDGFRa$^+$ mesodermal precursors are capable of forming mesenchymoangiobalst (MB) and hemangioblast (HB) colonies. Further mesodermal commitment to endothelial and haematopoietic cells requires downregulation of PDGFRa and primitive streak genes and along with upregulation of *TAL1*, *GATA2*, and *ETV2* genes leading to the commitment of hematovascular mesodermal precursors (HVMPs). **(B)**Upon gaining VE-cadherin expression, cells gradually acquire an endothelial or haematopoietic cell morphology and gene expression profile. The HE population can be discriminated from non-HE and early haematopoietic cells based on the lack of CD43 and CD73 expression. Further enrichment of the HE population can be achieved by low expression DLL4. Within the non-HE populations cells divide based on CXCR4$^+$ expression. The earliest haematopoietic progenitors emerge within the VE-cadherin+ population and have a CD43low phenotype. These cells are termed angiogenic haematopoietic progenitors (AHPs) as they retain endothelial potential. **(C)**Advanced haematopoietic development occurs through the upregulation of CD43 expression in which distinct subsets occur such as erythro-megakaryocytic progenitors (EMkP) and multipotent myelolymphoid progenitors. Figure adapted from (Slukvin, 2016), Created with BioRender.com

The differentiation of human PSCs towards HSPCs has been accomplished through two main methods; feeder dependent co-culture conditions or chemically defined culture conditions, that incorporate specific signalling pathway agonists and antagonists. Once HSPC have been generated these cells can be further differentiated into specific lineages including erythrocytes, megakaryocytes, platelets, macrophages and lymphoid cells.

3.1.2.1 Co-culture haematopoietic culture system

Various different feeder layers have been investigated for their ability to initiate haematopoietic fate in hPSCs such as; murine BM stromal cell lines S17 and OP9 or stromal cells derived from mid-gestational embryonic sites such as FL, YS and AGM (Qiu *et al.*, 2005; Vodyanik *et al.*, 2005; Ledran *et al.*, 2008). The most routinely used feeder layer is the OP9 cell line, as HSPCs can be generated from hPSC's in large numbers and in a short time frame, in comparison to other techniques, without the requirement of additional cytokines (Vodyanik *et al.*, 2005). A detailed protocol using the OP9 system has been established by the laboratory of Dr Igor Slukvin for the generation of HPSC's and mature myeloid cells, comprising three main steps; initiation of haematopoietic differentiation through co-culture of PSC's with OP9 BM cells, short-term expansion of multipotent myeloid progenitors using high dose GM-CSF (based on $lin^-CD34^+CD43^+CD45^+$ phenotype) and their directed differentiation of into specific myeloid lineages (Choi, Vodyanik and Slukvin, 2011). Furthermore, HSPCs generated from the OP9 co-cultures can be further differentiated into the lymphoid lineage through isolation of $CD34^+$ HSPCs from day 10 OP9 co-cultures and re-culture in a second differentiation step depending on the lineage of choice. B and NK cells can be produced by seeding HSPCs onto MS-5 stromal cells in the presence of SCF, Flt3-L and IL-7 (Vodyanik *et al.*, 2005) along with T-cells which are achieved by transfer of day 10 HSPCs onto OP9-DL1 feeders that ectopically express the Notch ligand, Delta-like 1 (Dll1) with further supplementation of FLT3-L, IL-7 and SCF (Schmitt and Zúñiga-Pflücker, 2002; Timmermans *et al.*, 2009). Mature myelomonocytic cells can also be produced with additional cytokine supplementation GM-CSF, interleukin 4 (IL-4) and tumor necrosis factor-α (TNF-α) lead to the generation of DCs; GM-CSF, transforming growth factor-β1 (TGF-β1) and TNF-α for Langerhans cells (LCs); M-CSF and IL-1β for macrophage production; G-CSF for neutrophils and IL-3 and IL-5 for eosinophils (Choi, Vodyanik and Slukvin, 2011). Although the OP9 system is very efficient at generating both HE and HPSC populations, the cultures are very sensitive to variations in serum quality, feeder layer maintenance and the size of PSC clusters used to initiate differentiation (Choi, Vodyanik and Slukvin, 2011). With these factors in mind, the study of signalling pathways involved at specific transitional phases during the differentiation process is difficult using this system due to the variability between experiments. Additionally, due to the use of xenogeneic

material throughout this protocol, haematopoietic cells generated are not able to be used for clinical applications (Uenishi *et al.*, 2014).

3.1.2.2 Chemically defined haematopoietic culture systems

Alternatively, haematopoietic differentiation can be carried out using a cocktail of cytokines, morphogens and small molecules in order to modulate signalling pathways at the different commitment stages in a serum-free media (Ramos-Mejia *et al.*, 2010; Uenishi *et al.*, 2014; Slukvin, 2016; Alsayegh *et al.*, 2019). Protocols ustilise Wnt agonists and bone morphogenetic protein 4 (BMP4) to induce mesoderm specification (Chadwick *et al.*, 2003; Yu *et al.*, 2011), Fibroblast Growth Factor 2 (FGF2) and Vascular Endothelial Growth Factor (VEGF) to promote hematovascular specification and endothelial-haematopoietic transition (Cerdan, Rouleau and Bhatia, 2004; Pick *et al.*, 2007) and an array of haematopoietic cytokines (SCF, FLT3L, IL-3, IL-6, G-CSF) are used to promote final haematopoietic commitment as these factors have been shown to be necessary during blood cell formation during embryogenesis (Slukvin, 2013b). These cytokine cocktails can also be modified depending on the lineage specification required and as described previously can be differentiated into lymphoid cells with the addition of modified stomal feeder layers (Schmitt and Zúñiga-Pflücker, 2002; Timmermans *et al.*, 2009; Choi, Vodyanik and Slukvin, 2011; Sturgeon *et al.*, 2014). These cultures can be initiated in two formats; two-dimensional *in vitro* aggregation of hPSC's into clusters termed 'Embryoid bodies' (EB's) that allow spontaneous and random differentiation into all three germ layers; ectoderm, endoderm and mesoderm or alternatively, monolayer cultures can be used that utilise extracellular matrix-coated plates, which has shown to have less variability (Ng *et al.*, 2005; Wang *et al.*, 2006; Real *et al.*, 2012; Castaño *et al.*, 2019).

3.2 Aims and objectives

The specific aims of this chapter are:

1. To characterise human iPS cell lines (KOLF2, and EPI) for downstream assays:
    - Assess state of pluripotency of individual lines
    - Assess genomic stability of individual lines

2. Optimise iPSC haematopoietic differentiation using established protocols.
    - Directed differentiation to haematopoietic progenitor cells (HSPCs):
        - OP9 coculture protocol
        - STEMdiff™ Kit (StemCell technologies)

    - Assess functionality of iPS derived HPC's using the colony forming cell (CFC)

3.3 Results

3.3.1 Characterisation of parental human iPS cell lines.

Due to the high variability of somatic cell type and reprogramming methods, iPSC lines were validated for their suitability for down-stream gene editing experiments. The human Episomal (EPI) iPS cell line was derived from $CD34^+$ cord blood cells using a three-plasmid seven factor EBNA-based episomal system (SOKMNLT; SOX2, OCT4 (POU5F1), KLF4, MYC, NANOG, LIN28, and SV40L T antigen) (Burridge *et al.*, 2011). The human iPS cell line KOLF2 was derived from fibroblast cells from a 55-year-old male using the CytoTune-iPS Sendai Reprogramming system containing four Sendai vector (SeV) based reprogramming vectors, each expressing one of the four Yamanaka factors (Oct4, Sox2, Klf4 and c-Myc).

3.3.1.1 Assessment of pluripotency.

First, human iPSC lines were examined for expression of cell-surface antigens and developmentally regulated genes widely used as markers of human ESCs in order to determine their state of pluripotency (Adewumi et al., 2007). On initial examination, during maintenance cultures both lines maintained typical pluripotent morphology, forming tightly packed colonies with defined borders and a high nucleus-to-cytoplasm ratio. Immunofluorescence analysis revealed that all lines expressed pluripotency specific antigens SOX2 and OCT3/4 in complementary nuclear patterns, consistent with DAPI nuclear staining along with cell surface antigen Stage-specific embryonic antigen-4 (SSEA-4) which were also complimentary to the human ESC-H7 line used as a positive control **(Figure 3.2 A and Figure 3.3 A)**. This was further confirmed by quantitative flow cytometry analysis, which validated 80-90% of the total population expressed SOX2, OCT3/4 and TRA-1-60 protein **(Figure 3.2 B and Figure 3.3 B)**. Collectively, this data demonstrates that the human iPS cell lines; EPI and KOLF2 carry the hallmarks of pluripotency.

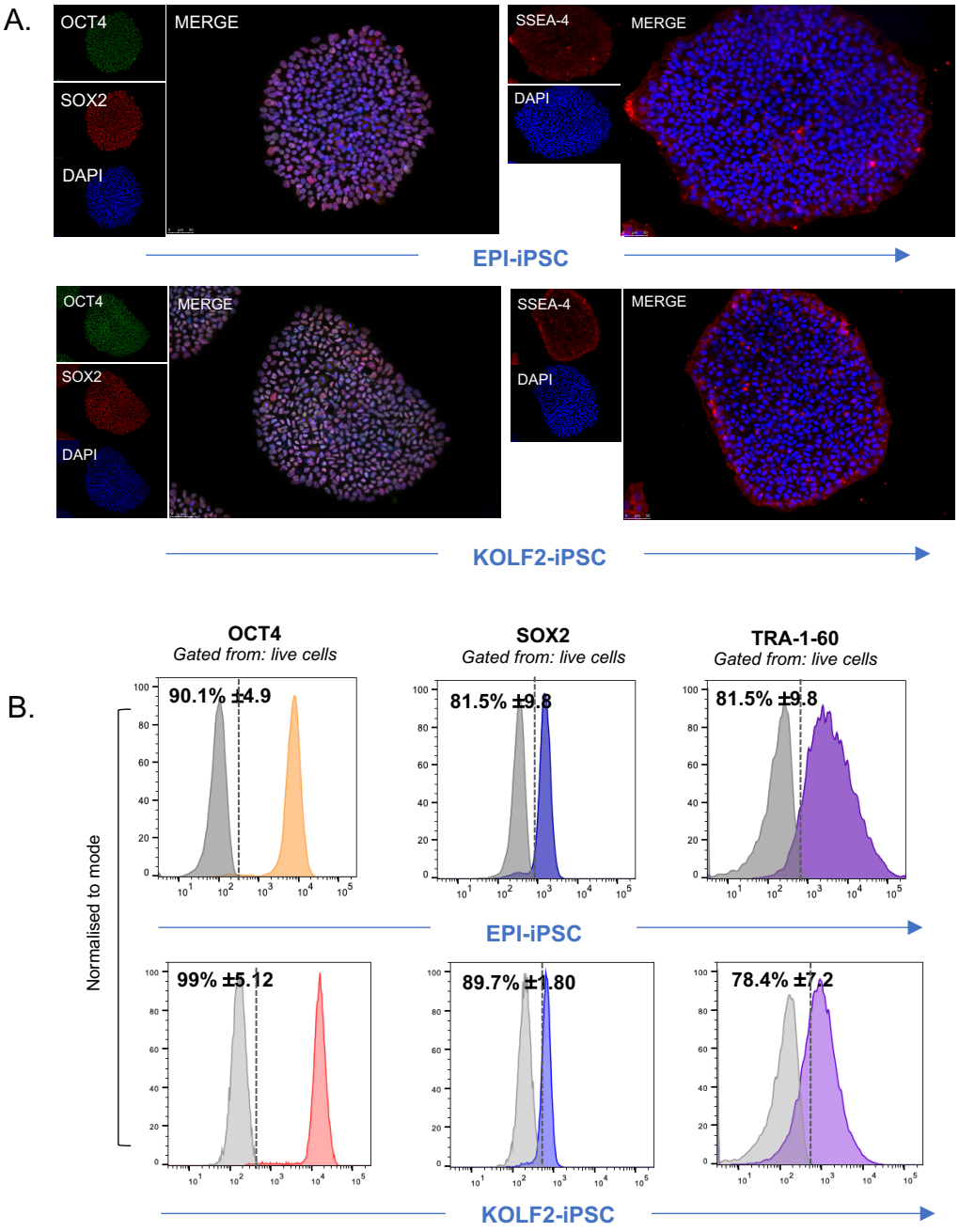


Figure 3.2 Human iPS cell lines EPI and KOLF2 carry the hallmarks of pluripotency. (A) Representative immunofluorescent images of human iPS colonies from EPI and KOLF2 lines stained with antibodies against OCT4, SOX2 and SSEA-4 and nuclei stained with DAPI. The human H7-ESC and human 293T cell lines were used as a positive and negative control (n=2). **(B)** Representative histogram plots showing percentage of cells expressing intracellular OCT4 and SOX2 and extracellular TRA-1-60, compared to isotype controls (grey). Data shown median values from 4 independent repeats with SEM.

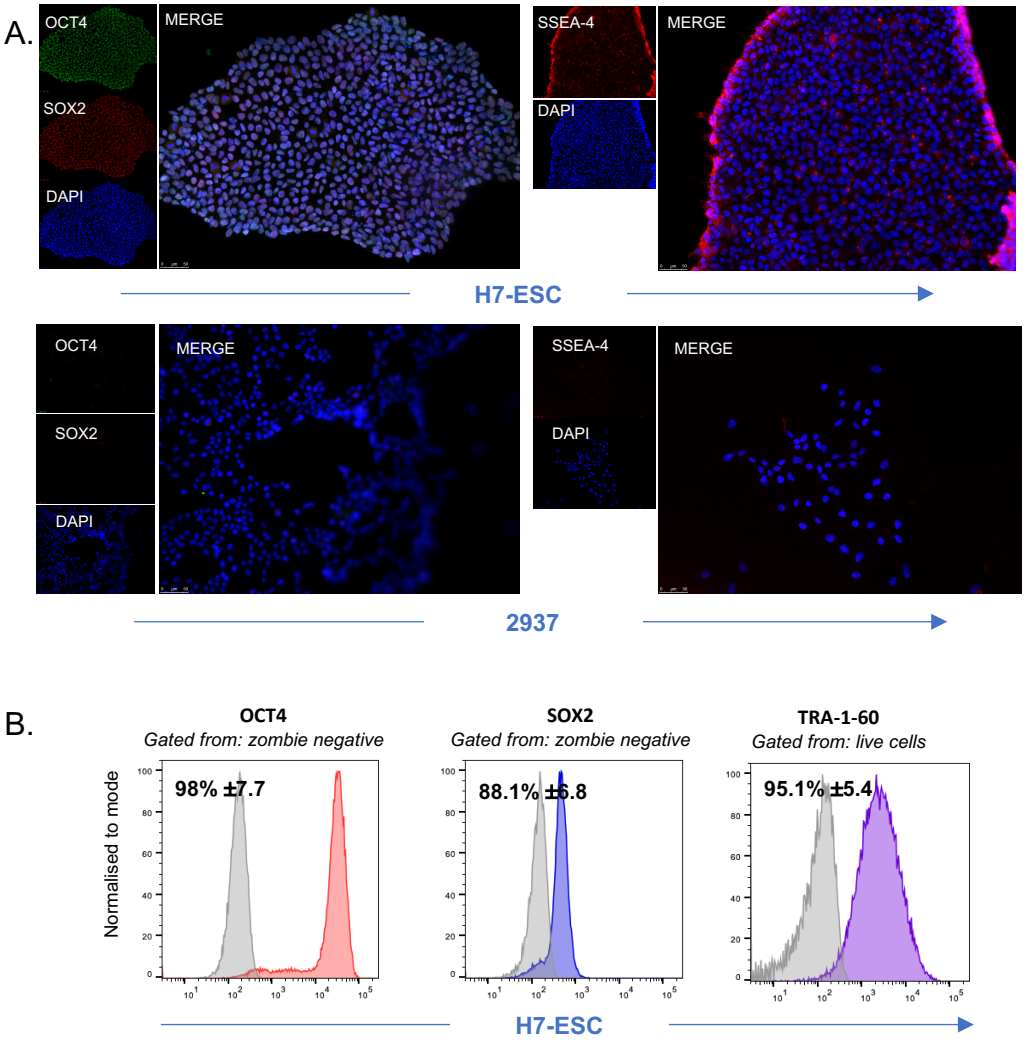


Figure 3.3: Human cell lines used as positive and negative controls for pluripotency analysis (A) Representative immunofluorescent images of human H7-ESC colonies and human 293T cells stained with antibodies against OCT4, SOX2 and SSEA-4 and nuclei stained with DAPI. (n=2). **(B)** Representative histogram plots of human H7-ESC cells showing expression of intracellular OCT4 and SOX2 and extracellular TRA-1-60, compared to isotype controls (grey). Data shown median values from 3 independent repeats with SEM.

3.3.1.2 Assessing genomic stability of human iPS lines.

Next, genomic stability of both iPSC lines were investigated, as it has been well documented that prolonged passaging and improper cell maintenance can result in culture adaptation and genetic instability in pluripotent stem cells (PSCs) (Caisander et al., 2006; Baker et al., 2007; Grandela and Wolvetang, 2007). Karyotyping was carried out to screen for abnormalities such as monosomy and trisomy of chromosomes acquired through extended passaging. Cells were treated with Nocodazole to arrest mitotic division at the metaphase stage, incubated in a hypotonic solution of 0.4% KCL and dropped onto slides to perform chromosome counts. Both

lines presented with a euploid karyotype represented by a stable chromosome count of 46 in >70% of metaphase spreads analysed **(Figure 3.4)**.

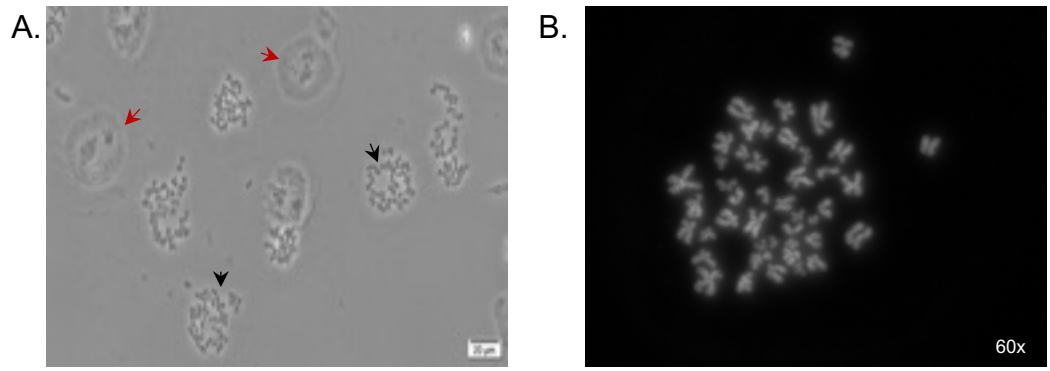


Figure 3.4: Human PSC lines present with a euploid karyotype according to metaphase counts. Cells were arrested at the metaphase stage of division using Nacodazole (0.1ug/mL), incubated in a hypotonic buffer (0.4% KCL with HEPES) and spread onto slides in order to expose chromosomes from the nuclei. **(A)** Representative metaphase spreads from the EPI-iPS cell line. Arrows indicate successful metaphase spreads suitable for counting (black) and intact nuclei (red). **(B)** Chromosomes were stained with DAPI DNA stain and individual metaphase spreads were imaged under a light microscope (Olympus BX43) and individual chromosomes counted to assess cell line stability.

3.3.2 Optimising human iPS haematopoietic differentiation

It has been well documented that there is great variation between human iPS cells in their potential to differentiate into specific lineages. This variation can be attributed to differences in the genetic backgrounds of individual donors (Féraud *et al.*, 2016), the reprogramming process (both from the reprogramming factors used and the different delivery methods), residual DNA methylation signatures from the reprogramming process and the retention of 'epigenetic memory' from the tissue of origin (Noguchi, Miyagi-Shiohira and Nakashima, 2018). It is therefore essential to assess the quality of any human iPS lines in terms of their capability to differentiate into the lineage of interest.

3.3.2.1 Generating HSPCs using the OP9 coculture system

To determine the capacity of human iPS cell lines EPI and KOLF2, to differentiate towards the haematopoietic lineage, the co-culture approach was used in accordance with the Vodyanik and Slukvin model (Vodyanik and Slukvin, 2007). This protocol has widely been used in the literature to characterise and isolate haematopoietic and endothelial progenitors based on expression of CD43, CD31 and CD34 (Vodyanik *et al.*, 2005; Choi *et al.*, 2009; Chicha *et al.*, 2011). Murine OP9 stromal cells were cultured to confluency and subsequentially conditioned

for 4-6 days. HiPS cells were seeded at defined densities as clusters onto the OP9 stromal layer in haematopoietic differentiation media containing haematopoietic cytokines. The media was replaced at day one, followed by half a media change on days four and six. At day eight the cultures were dissociated and immunophenotyped for endothelial and haematopoietic potential based on surface marker expression **(Figure 3.5)**.

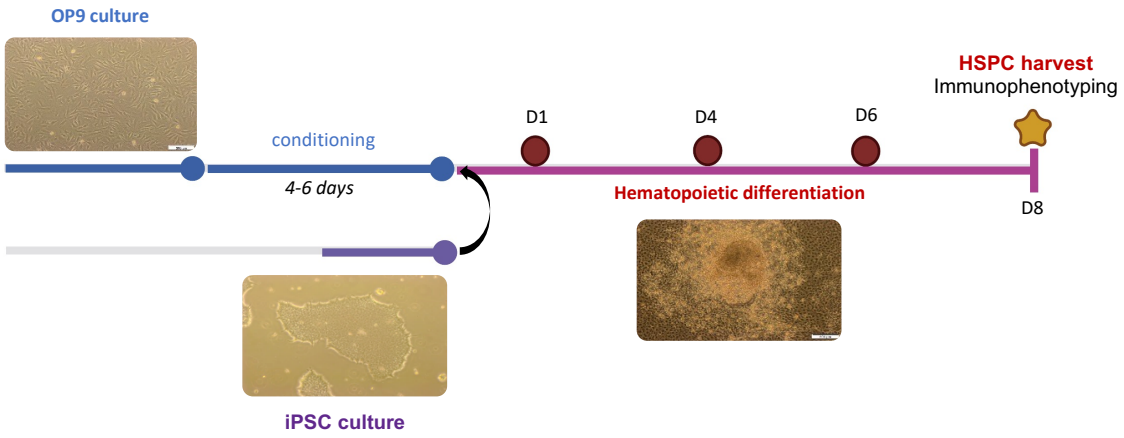


Figure 3.5: Human pluripotent stem cell haematopoietic differentiation using the OP9 co-culture system. Schematic diagram of the protocol used; human iPS cells were maintained in the E8 flex system before being seeded onto conditioned OP9 stromal cells for eight days with half media changes on days 4,5 and 6 (red circles). Cells were harvested on day 8, dissociated and immunophenotyped by flow cytometry.

Current protocols in the literature using this OP9 co-culture method, state that during the initiation of haematopoietic differentiation cultures, human iPSC are seeded onto OP9 stromal cells that have undergone a conditioning phase post confluency for a period of 4-8 days (Vodyanik and Slukvin, 2007; Choi, Vodyanik and Slukvin, 2011). To define the optimum culture time for OP9 conditioning prior to seeding of human iPS cells, we investigated the HPC output of cultures initiated from OP9 cells conditioned from 4-6 days post confluency using the human iPS EPI line. Cells were phenotypically analysed after eight days by flow cytometry. The majority of the $CD34^+$ cells produced during the differentiation (independent of OP9 conditioning) expressed an endothelial phenotype marked by $CD31^+CD43^-$ surface markers (80-90%), with no clear population of haematopoietic progenitors being observed ($CD31^+CD43^+$) **(Figure 3.6)**. It has been documented that the efficiency of haematopoietic differentiation is significantly affected by the density of iPSCs seeded onto the OP9 feeder layer (Vodyanik and Slukvin, 2007; Choi, Vodyanik and Slukvin, 2011). Nevertheless, a range of densities have been described within these protocols. Therefore, we attempted to optimise the seeding density used to initiate differentiation cultures. We investigated the effect of two iPSC densities ($1x10^6$cells to $1.5x10^6$cell/T75cm) on HPC output using the EPI iPSC cell line,

to assess whether this could explain the lack of commitment into the haematopoietic lineage observed in the initial experiments. Though we observed an increase in the total number of $CD34^+$ cells **(Figure 3.7 A)** along with a reduction in cells with an endothelial immunophenotype ($CD31^+CD43^-$), we failed to detect a robust HPC population with $CD34^+CD31^+CD43^+$ immunophenotype **(Figure 3.7 B)**.

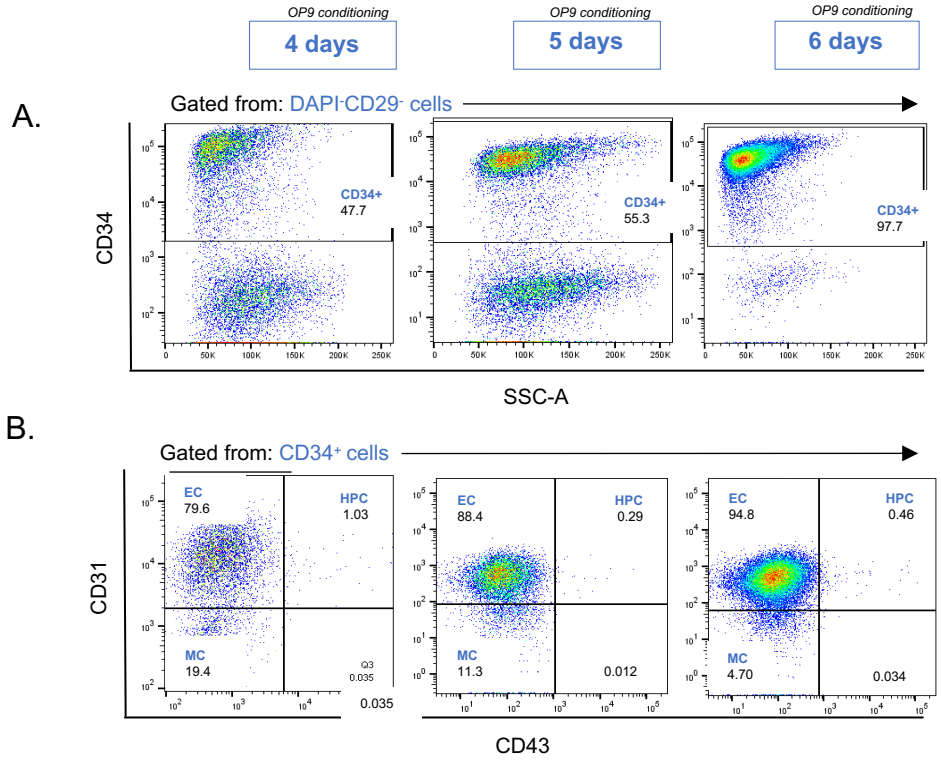


Figure 3.6: Phenotypic analysis of EPI-iPSC initiated on OP9 co-cultures with varying conditioning phases. EPI-iPS cells were used to initiate differentiation cultures on OP9 cells that had undergone conditioning for 4, 5 or 6 days post confluency. Representative flow cytometry end point analysis at day 8 from $CD34^+$ enriched population **(A)** frequency of $CD34^+$ population gated from live human cells ($DAPI^-$ $CD29^-$). **(B)** The CD34+ population was further subdivided into $CD34^+CD31^+CD43^-$ endothelial cells (EC), $CD34^+CD31^+CD43^+$ haematopoietic progenitors (HPC) and $CD34^+CD31^-CD43^-$ mesenchymal cells (MC). Numbers in each quadrant indicate the mean of two independent repeats (n=2) as a percentage of the previous gate.

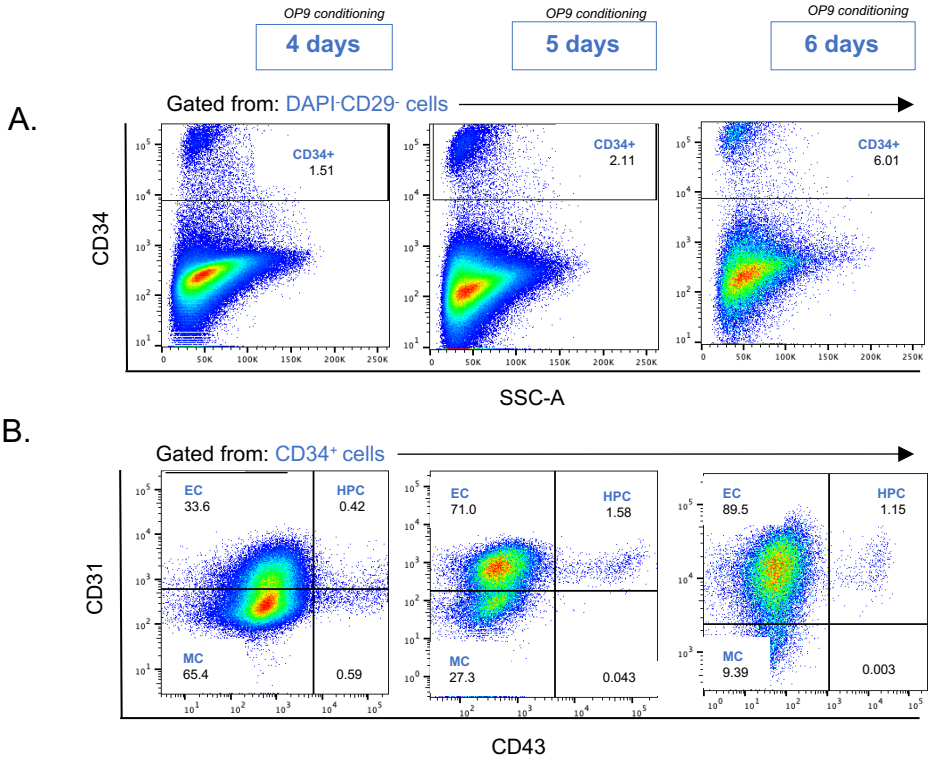


Figure 3.7: Phenotypic analysis of EPI-iPSC seeded at a higher density to initiate OP9 co-cultures varying OP9 conditioning phases. An increased cell density (1.5x10^6) of EPI-iPS cells were used to initiate differentiation cultures on OP9 cells that had undergone conditioning for 4, 5 or 6 days post confluency. Representative flow cytometry end point analysis at day 8 **(A)** 1x10^6 cells were separated from total harvested cells and analysed for frequency of CD34$^+$ population gated from live human cells (DAPI$^-$ CD29$^-$) **(B)** the remaining total cells were enriched for CD34$^+$ expressing cells and immunophenotyped for CD34$^+$CD31$^+$CD43$^-$ endothelial cells (EC), CD34$^+$CD31$^+$CD43$^+$ haematopoietic progenitors (HPC) and CD34$^+$CD31$^-$CD43$^-$ mesenchymal cells (MC). Numbers in each quadrant indicate the mean of two independent repeats (n=2) as a percentage of the previous gate.

Lastly, we repeated haematopoietic differentiation at the same hiPS seeding density using OP9 stromal cells conditioned for 5 or 6 days, using the human iPS-KOLF2 cell line, reported in the literature to have good haematopoietic differentiation capacity. We observed an increase in the proportion of CD34$^+$ cells at day 5 (6.68%) and day 6 (8.80%) when compared with the hiPS EPI line **(Figure 3.8 A)**. However, the frequencies of CD34$^+$CD31$^+$CD43$^-$ endothelial cells (6.01% at day 5 and 14% at day 6), CD34$^+$CD31$^+$CD43$^+$ haematopoietic progenitor cells (1.20% at day 5 and 0.8% at day 6) and CD34$^+$CD45$^+$ were decreased **(Figure 3.8 A and B)** in comparison with the hiPS-EPI line.

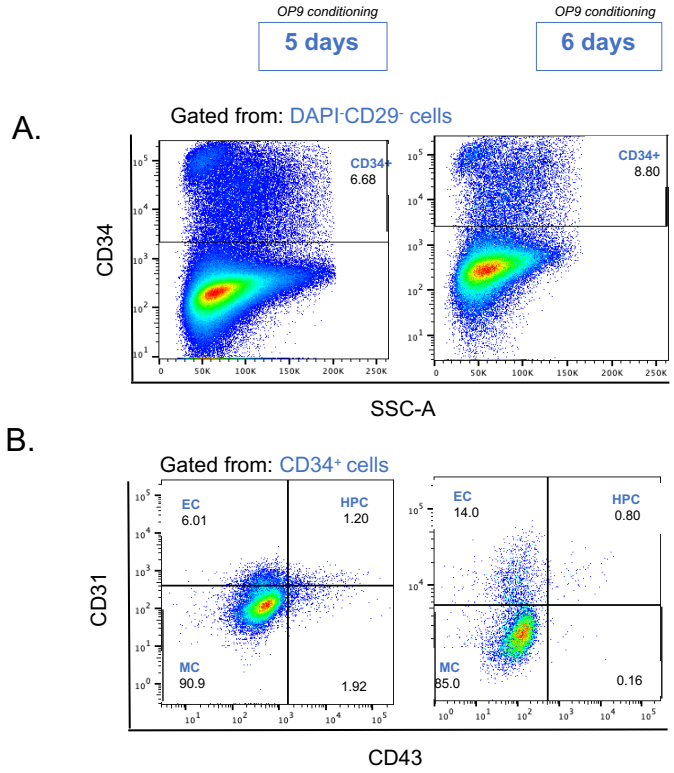


Figure 3.8: Phenotypic analysis of KOLF2-iPSC haematopoietic differentiation using the OP9 co-culture system with varying conditioning phases. $1.5x10^6$ KOLF2-iPS cells were used to initiate differentiation cultures on OP9 cells that had undergone conditioning for 5 or 6 days post confluency. Representative flow cytometry end point analysis at day 8 **(A)** $1x10^6$ cells were separated from total harvested cells and analysed for frequency of $CD34^+$ population gated from live human cells ($DAPI^-$ $CD29^-$) **(B)**the remaining total cells were enriched for $CD34^+$ expressing cells and immunophenotyped for $CD34^+CD31^+CD43^-$ endothelial cells (EC), $CD34^+CD31^+CD43^+$ haematopoietic progenitors (HPC) and $CD34^+CD31^-CD43^-$ mesenchymal cells (MC). Numbers in each quadrant indicate the mean of two independent repeats (n=2) as a percentage of the previous gate.

3.3.2.2 Generating HSPCs using STEMdiff™ system

To assess whether the low efficiencies observed in producing a $CD34^+CD31^+CD43^+$ HPC populations from both iPS cell lines were a reflection of their limited differentiation capacity into the haematopoietic lineage or due to ineffective optimisation of the OP9 co-culture system, we re-assessed their potential using an alternative differentiation system fully optimised by StemCell technologies- the STEMdiff™ haematopoietic differentiation kit. This protocol involves the differentiation of iPSCs in a feeder-free system, using a fully defined media in order to overcome the low differentiation efficiencies and high variability from undefined components such as feeder cell lines (as observed from initial experiments) using a standardized protocol. iPSCs were dissociated into small aggregates, seeded onto Matrigel coated culture plates, induced towards mesoderm commitment followed by subsequent commitment towards haematopoietic progenitor **(Figure 3.9)**.

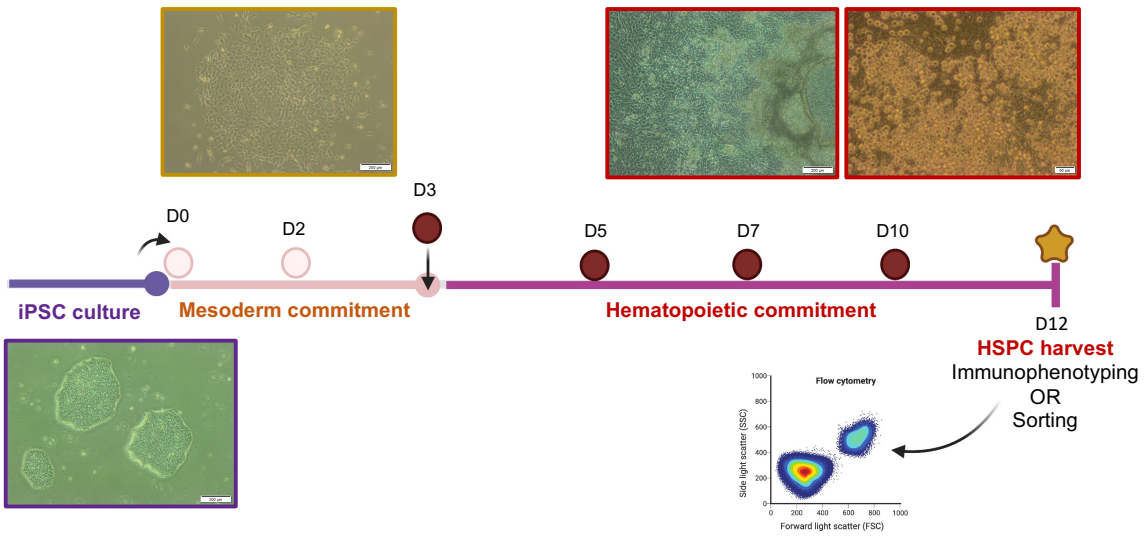


Figure 3.9: Human iPSC haematopoietic differentiation using the STEMdiff™ culture. Schematic diagram of the two-step haematopoietic differentiation protocol; human iPSCs were maintained in E8 Flex media, dissociated into small clusters and seeding onto Matrigel coated culture dishes. Mesoderm commitment was induced in the first three days using a cocktail of pre-defined cytokines (pink circles represent half media changes post initial media change), followed by haematopoietic progenitor commitment again using pre-defined cytokines (red circles represent half media changes post initial media change). At day twelve the cultures were harvested and either immunophenotyped or sorted to test the functionality of HPCs produced by plating CFC assays.

Using the STEMdiff™ system we differentiated the EPI-iPSC line and observed growth of mesoderm-like colonies with endothelial cell characteristics from the seeded iPS cell during the initial commitment phase of the differentiation (days 1-3). Upon commitment of the haematopoietic lineage, from as early as day eight, clusters of cells could clearly be observed

emerging at the periphery of mesoderm-like colonies. By day 10 and 12 these cells filled the supernatant and exhibited typical HPC morphology; characterised as round cells that float freely in suspension **(Figure 3.10 A)**. The differentiation culture was harvested at day 12 and immunophenotyped where around 30% of $CD34^+$ cells were observed with the $CD31^+CD43^+$ phenotype typical of multipotent progenitors (**Figure 3.10 B)**.

As haematopoietic differentiation is a continuous process in which multipotent progenitors develop within a specific time frame highly dependent on the culture conditions, we next aimed to establish the kinetic profile of haematopoietic commitment, which has yet to be defined in the STEMdiff™ system. To achieve this, we harvested and immunophenotyped differentiation cultures every 2 days post mesoderm commitment (day 4), as cell number prior to this time point were too low for sufficient analysis by flow cytometry **(Figure 3.11)**. As the cultures developed, we observed an increase in the frequency of $CD34^+$ cells from 6.6% to 56%. Within the $CD34^+$ population we observed an increase in endothelial cells expressing $CD34^+CD31^+CD43^-$ during early time points, which peaked at day 8 accounting for 51% of the total cells in the culture which declined to 35% at the day 12. Cells expressing $CD34^+CD31^+CD43^+$ which have broader lymphomyeloid differentiation potential were detected at low level as early as day 4 and steadily increased over the culture period to 49% of the final cell population. Finally, a $CD34^+CD31^+CD43^+CD45^+$ population highly enriched in myeloid progenitors appeared between day 10 and 12 accounting for 26% of the culture **(Figure 3.12 A)**. To assess the functionality of HPCs generated at day 12, $CD34^+CD31^+CD43^+$ cells were sorted into Colony forming Unit (CFU) assays where cells are cultured in methycellulose media enriched with cytokines that support haematopoietic progenitors with clonogenic potential. These HPCs displayed high CFU potential with multilineage haematopoietic colony types, which was predominantly myeloid growth in-line with our results showing approximately half of these HPCs expressed CD45 expression marking their specification towards the myeloid lineage **(Figure 3.12 B)**.

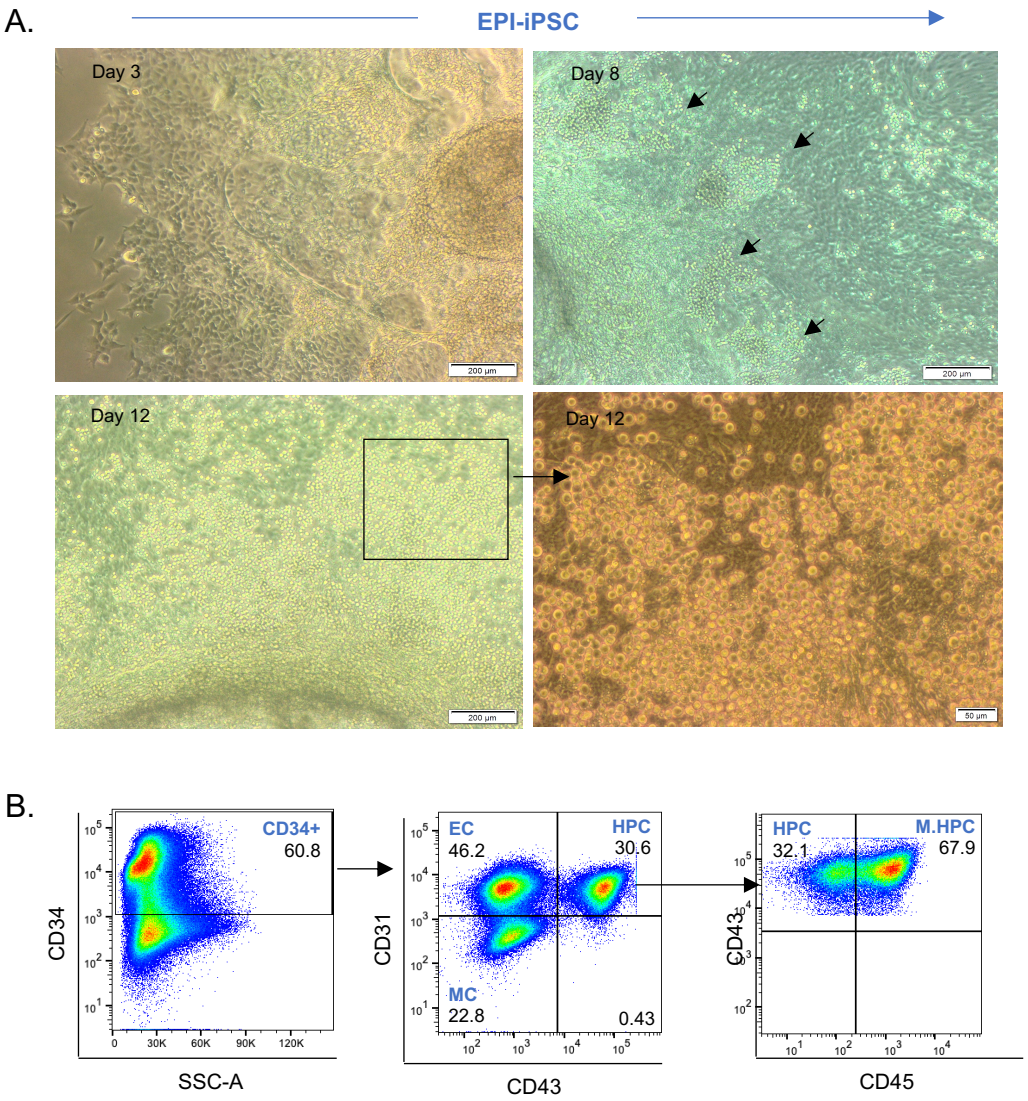


Figure 3.10: The human EPI-iPS cell line efficiently differentiates into HPCs using the STEMdiff™ culture system. (A) Representative images of the haematopoietic differentiation from EPI-iPSC line at days 4,8 and 12; mesoderm-like colonies emerged from seeded iPS cells at day 3 of culture, haematopoietic cells start emerged at periphery of these colonies at day 8, by day 12 an abundance of uniform haematopoietic cells are observed throughout the culture which can be easily harvested from the culture supernatant. **(B)** Representative flow cytometry analysis of cells harvested from the culture at day 12: $CD34^+$ cells were gated from live cells and then subdivided into mesenchymal cells (MC), endothelial cells (EC) and haematopoietic progenitor cells (HPC), these HPCs mature and become $CD45^+$ (M.HPC) (left to right).

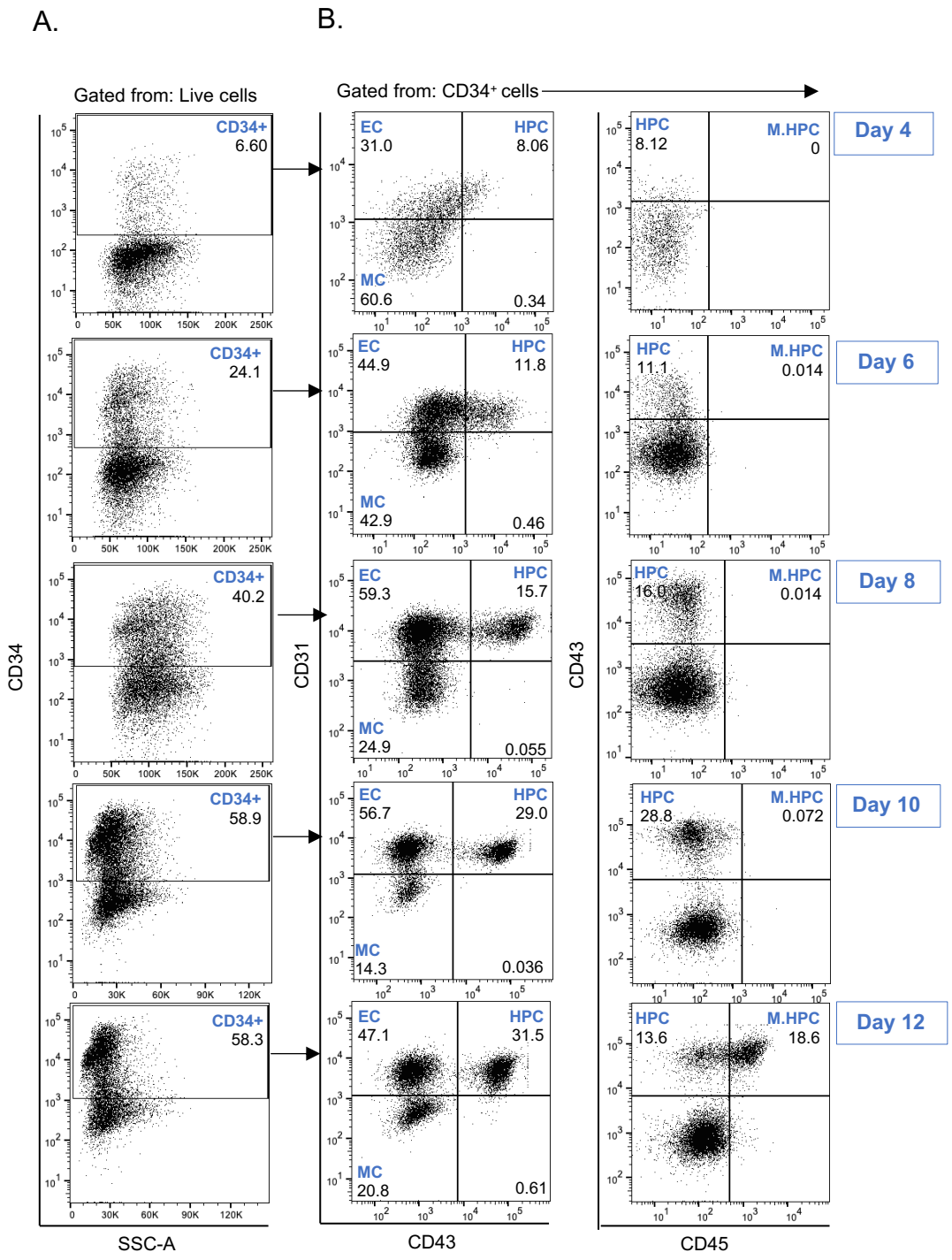


Figure 3.11: Kinetic analysis of haematopoietic development in human iPS-EPI line. Representative flow cytometry dot plots displaying **(A)** $CD34^+$ cells, gated from DAPI- live cells and **(B)** further dissection of the $CD34^+$ compartment based on the expression of surface markers $CD31^+CD43^-$ endothelial cells (EC), $CD31^+CD43^+$ haematopoietic progenitor cells (HPC), $CD31^-CD43^-$ mesenchymal cells (MC) and $CD31^+CD43^+CD45^+$ myeloid enriched HPCs (N=2).

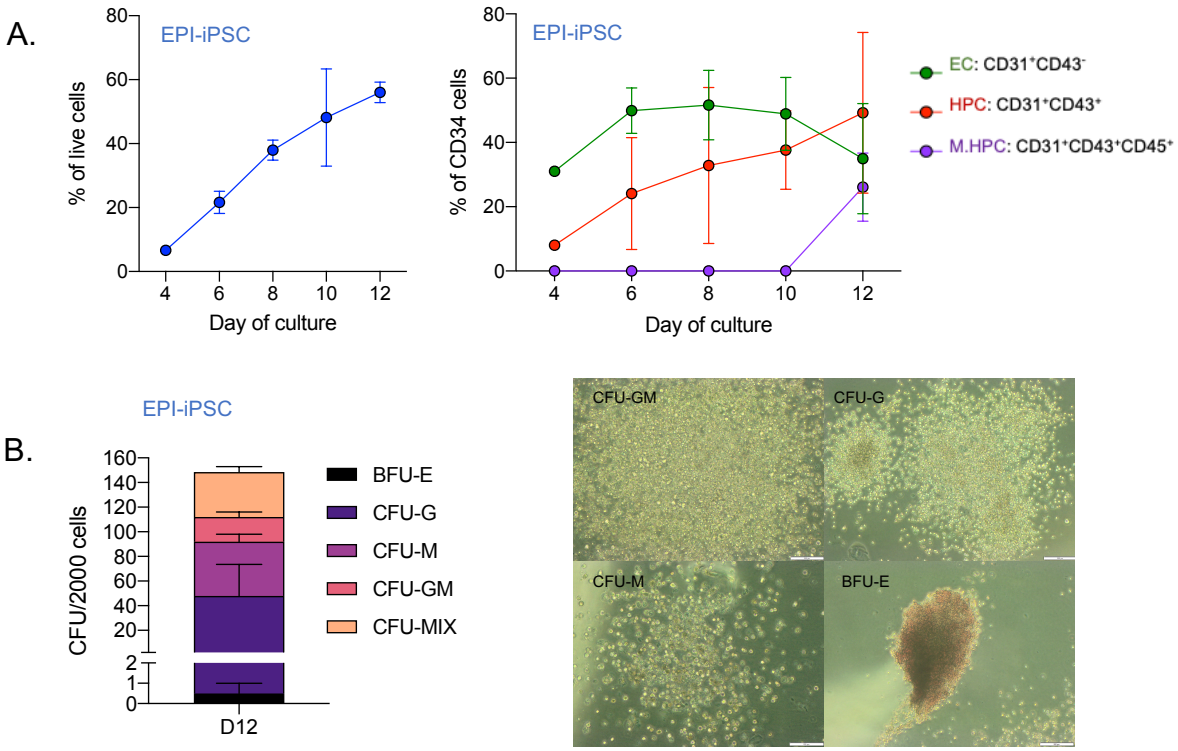


Figure 3.12: The EPI-iPSC line efficiently produces HPCs and has high CFU potential with multilineage capacity. Kinetic analysis of haematopoietic differentiation using the EPI and KOLF2 iPSC lines **(A)**Frequency of $CD34^+$ cells gated from live cells and frequency of $CD31^+CD43^-$ endothelial cell (EC) population, $CD31^+CD43^+$ haematopoietic progenitor cell (HPC) population and $CD31^+CD43^+CD45^+$ mature HPC population (N=2). **(B)**Frequency of Colony-forming unit potential and distribution from 2,000 sorted $CD34^+CD31^+CD43^+$ HPCs and representative images of CFU type from EPI-iPSC line (N=2). Assays were routinely scored at day 14. Data shown are SEM.

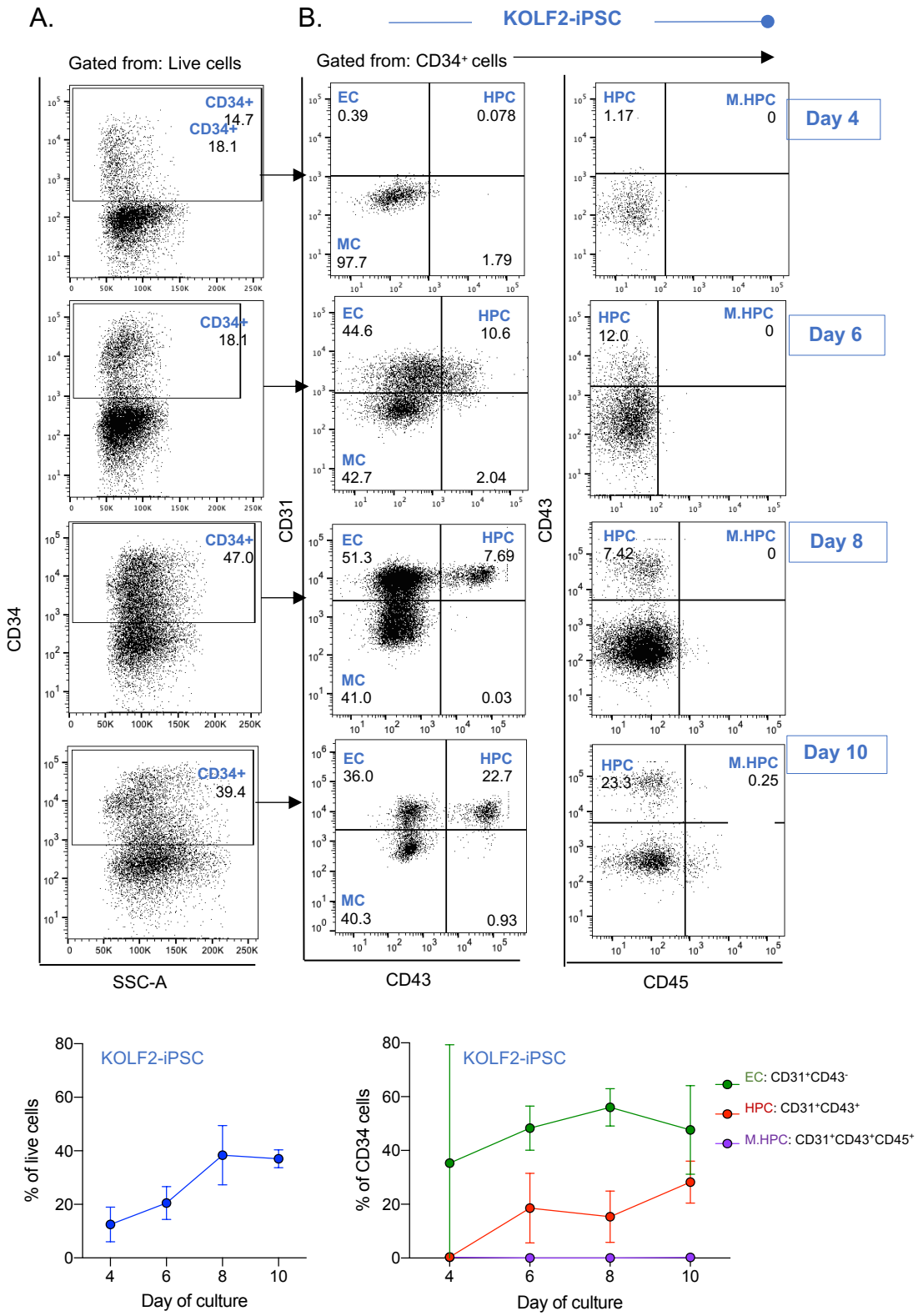


Figure 3.13: Kinetic analysis of haematopoietic development in human iPS-KOLF2 line. Representative flow cytometry dot plots displaying (A) $CD34^+$ cells, gated from DAPI- live cells and (B) further dissection of the $CD34^+$ compartment based on the expression of surface markers $CD31^+CD43^-$ endothelial cells (EC), $CD31^+CD43^+$ haematopoietic progenitor cells (HPC), $CD31^-CD43^-$ mesenchymal cells (MC) and $CD31^+CD43^+CD45^+$ myeloid enriched HPCs (C) Kinetic analysis summary (N=3).

Finally, with the knowledge that this haematopoietic differentiation culture system was effective in producing high yields of HPCs in our hands we repeated the kinetic analysis of our second iPS cell line KOLF2 to assess the differentiation potential. We observed a similar increase in $CD34^+$ cells as the differentiation culture proceeded along with acquisition of cells with endothelial phenotype which followed the same trajectory as the EPI-iPSC line. However, HPCs expressing the $CD34^+CD31^+CD43^+$ phenotype was detected at a slightly later time point (day 6) and the abundance endothelial cells was higher (47%) at day 12 with a lower abundance of HPCS (28%) in comparison to the EPI line **(Figure 3.13)**.

Taken together, our results show that the EPI and KOLF2 iPSC lines are capable of differentiating into haematopoietic progenitor cells as observed from our phenotypic analysis and CFU assays. Moreover, as indicated in the literature we show that there is variability between lines, with the EPI line being more efficient in producing a high abundance of $CD34^+CD31^+CD43^+$ HPCs.

3.4 Discussion

In this chapter we aimed to characterise two human iPS cell lines EPI and KOLF2 for their suitability as parental lines for downstream modelling experiments by determining their state of pluripotency, genomic stability and their ability to differentiate into the haematopoietic lineage. We showed that, both lines carried the hallmarks of pluripotency, displayed a normal diploid karyotype of 46 and were capable of differentiating into haematopoietic progenitors (HPC) with multilineage CFU potential.

When optimising the haematopoietic differentiation of these lines we faced many issues. In our first attempts we employed the OP9 co-culture method that has been successfully used by many groups to differentiate both murine and human pluripotent stem cells into multilineage haematopoietic progenitors (Slukvin *et al.*, 2006; Vodyanik and Slukvin, 2007; Uenishi *et al.*, 2014; Kumar *et al.*, 2020). This system requires very precise parameters in order for successful HPC output to be achieved which are highlighted in the published methods (Choi, Vodyanik and Slukvin, 2011). These include; ensuring OP9 feeder layers are grown at high densities whilst minimising spontaneous adipogenic differentiation, ensuring PSC are in an optimal dividing stage, have not been cultured for prolonged passages and finally used to initiate differentiation culture at specific densities (Vodyanik and Slukvin, 2007; Choi, Vodyanik and Slukvin, 2011). Nevertheless, we highlighted two main aspects of the protocol which are variable and effect differentiation output substantially. Initially we observed that seeding density can have a strong effect on the efficiency of the differentiation as we found the majority of PSC in our first experiments had developed through mesoderm and hematoendothelial commitment but were skewed towards endothelial commitment ($CD34^{+}CD31^{+}CD43^{-}$), resembling earlier time points of this differentiation system indicated by kinetic profiling conducted by Vodyanik, Thomson and Slukvin, 2006). Although we observed a small improvement by increasing the PSC density from 1 to 1.5 million cells which resulted in a small population with haematopoietic potential ($CD34^{+}CD31^{+}CD43^{+}$) this was not significant enough to produce a robust population. In line with these observations Kumar *et al.,* (2019) during OP9 co-culture witnessed substantial differences in the efficiency of PSC depending on their maintenance conditions, prior to differentiation. The authors stated that PSCs maintained on MEFs, outperformed their counterparts that were maintained in feeder-free conditions and that these discrepancies were attributed to bigger clumps being obtained when dissociating PSC from MEF, which then go on to differentiate better in OP9 co-culture, in comparison to the smaller clumps dissociated in feeder free cultures (Kumar, D'Souza and Thakur, 2019).

Subsequently, we addressed the variability of the length in which OP9 cells were initially conditioned in culture prior to PSC seeding. The literature states a window of 4-8 days in which OP9 feeder cells require conditioning before PSC seeding. However, our results clearly showed that there can be marked differences in haematopoietic output day by day- with the haematopoietic potential of $CD31^{+}CD43^{+}$ fraction of $CD34^{+}$ cells produced in the culture being increased >3 fold from day 4 conditioned OP9 cells to day 5 (Vodyanik and Slukvin, 2007; Choi, Vodyanik and Slukvin, 2011). Furthermore, it has been highlighted that the batch to batch variations found between FBS lots can affect the proliferation and differentiation of cells *in vitro* (Vodyanik and Slukvin, 2007). Groups have found that during OP9-coculture different lots of HyClone-define FBS provides relatively stable differentiation from pluripotent cells, whilst importantly sustaining OP9 maintenance without significant adipogenesis, which can impair differentiation in this system (Choi, Vodyanik and Slukvin, 2011). It is clear that the OP9 system requires much optimisation in order to produce robust numbers of HPCs from PSC cultures which must be done from lab to lab. Gathering our data and what has been reported in the literature, to limit variability observed in this system, future experiments using our EPI and KOLF2 lines would benefit from transitioning iPS cell lines onto MEF feeders or alternatively ensuring larger clumps from feeder free passaging were used to initiate differentiation cultures, using HyClone FBS and then assess the optimum time required for OP9 conditioning. Given the considerable difficulties we experienced in generating haematopoietic progenitors from the OP9 co-culture system we moved to a feeder-free haematopoietic differentiation system with less variables to control. Using the STEMdiff™ differentiation system both iPSC lines had the capacity to give rise to hemogenic progenitors which further differentiated into haematopoietic progenitor cells capable of CFU potential. This shows that these lines are capable of differentiating into the haematopoietic lineage and supports our theory that the OP9 co-culture requires further optimisation.

Taken together the data shows that these lines are suitable to use as parental lines in order to model GATA2 immunodeficiency syndromes using CRISPR/Cas9 editing and haematopoietic differentiation. As we aim to investigate the effects of specific *GATA2* mutations observed in patients which share the hallmark of immune dysfunction, the STEMdiff differentiation system would allow us to examine any developmental defects as a result of these mutations in HPCs. Additionally mature myelomonocytic cells such as neutrophils, eosinophils, DCs, Langerhans cells (LC), macrophages and osteoclasts can be obtained from an additional 5-19 days on OP9 feeders ectopically expressing the Notch ligand, Dll1 with appropriate cytokine combinations relevant for fate decisions of specific cell types (Slukvin *et al.*, 2006; Choi, Vodyanik and Slukvin, 2011). Furthermore, protocols have also been established for successful differentiation into lymphoid lineages such as NK, B and T cells

(Schmitt and Zúñiga-Pflücker, 2002; Vodyanik *et al.*, 2005; Timmermans *et al.*, 2009). As previously discussed in chapter one, many of these GATA2 patients present with significant cytopenias in one or more of these lineages therefore, being able to investigate any molecular changes that occur during these fate decisions would be invaluable for understanding how this syndrome develops from a premalignant stage.

Chapter 4:

Modelling patient specific GATA2 mutations using CRISPR/Cas9 technology

4.1 Introduction

4.1.1 Human iPSC disease modelling

Understanding the biological processes that underlie human disease is crucial in order to formulate strategies for their prevention and treatment. Due to the highly conserved nature of mammalian genomes, animal models such as mouse, rat and non-human primates have served as an invaluable tool for human disease modelling through the dissection of molecular mechanism during various developmental stages in a variety of cell types *in vivo.* With the advent of transgenesis and gene targeting, mouse models have become the most widely used model organism. However, humans and mice have a considerable amount of genetic, physiological and developmental differences, which is especially true during embryonic development (Avior, Sagi and Benvenisty, 2016). It is therefore more preferable to conduct research using human models through the culture of primary material in order to study disease aetiology at the molecular and cellular level. However, this too comes with its drawbacks as human tissues/cells are often limited, contain heterogeneous cell populations which are often not consistent between samples and are more difficult to culture (Bassett, 2017).

Since Yamanaka's ground breaking research, induced pluripotent stem (iPS) cell technology has revolutionised disease modelling through the reprogramming of normal and disease patient somatic cells towards a pluripotent stem cell state- almost identical to embryonic stem cells (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007). Human iPS cells have become widely used in place of other modelling strategies due three main advantages; they are normal primary cell lines, have the capacity for indefinite self-renewal and are most importantly they are a cell type from a very early stage of development and therefore can be used to model both developmental and differentiation defects, along with the temporal sequence of events in early stages of disease progression (Bassett, 2017).

In the last decade human iPS cells have been generated from several haematological malignancies including AML, MDS, MPN, and CML (Chao and Majeti, 2019; Turhan *et al.*, 2019). These studies have fulfilled three key criteria for effective iPS cell modelling from blood cancer; successful reprogramming from malignant cells to functional iPS cells assessed by typical *in vitro/vivo* characteristics, preservation of the underlying malignant somatic mutations and disease phenotype characterisation of malignant-derived iPS (Chao and Majeti, 2019). Studies recapitulating leukaemia using iPS cell reprogramming *in vivo* was not reported in previous years. However, recently two groups have successfully validated the ability of AML or MDS-derived iPS cells to recapitulate leukaemia when transplanted into immunodeficient mice (Chao *et al.*, 2017; Kotini *et al.*, 2017).

Though many successful iPS cell models have be reported from human myeloid malignancies, iPS generation from lymphoid malignancies seem to be hindered by the lack of tumour tissue available from cancers such as lymphoma. Whilst groups have attempted iPS reprogramming from these malignancies, there is yet to be any success (Muñoz-López *et al.*, 2016). Furthermore, many groups have reported difficulties in reprogramming patient-derived cells from myeloid malignancies as shown by one study reporting that from 16 AML patient samples only one iPS line was efficiently generated with leukemic specific mutations from one patient sample (Lee *et al.*, 2017). Other studies have reported that patient samples carrying specific mutations such as p53 and some complex karyotypes can have detrimental effects to the reprogramming process (Sarig *et al.*, 2010). Moreover, specific regions of the genome have been demonstrated to be refractory to the binding of Oct4, Sox2 and Klf4, which are vital in the molecular process of reprogramming and therefore patients carrying specific mutations in chromatin-modifying proteins, can reduce or abolish reprogramming altogether (Onder *et al.*, 2012). To circumvent some of these limitations many groups have alternatively generated human iPS cell lines carrying disease relevant mutations through genome editing techniques.

4.1.2 CRISPR-Cas genome editing tools

In the past decade the field of genome editing has made considerable progress with the advent of CRISPR/Cas technology. In their naturally occurring format, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins form the adaptive immune system in prokaryotes. The CRISPR-Cas9 adaptive immune system stores memory of previous infections by capturing pathogen nucleic acid sequences (‘spacer sequences’) in the CRISPR array. Upon re-infection these sequences are used to guide CRISPR-Cas proteins to seek and destroy pathogen DNA or RNA. The system allows the targeting of a large array of RNA and DNA sequences through the use of different spacer sequences within a guide RNA molecule that binds complimentary sequences adjacent to a suitable protospacer adjacent motif (PAM) in the target DNA. The requirement of a PAM sequence is to ensure that the genomic DNA encoding the guide RNA molecule is not targeted (Hille *et al.*, 2018; Anzalone, Koblan and Liu, 2020).

Since its initial discovery, the re-programmability of the CRISPR-Cas system and the variety of Cas effector proteins have been explored in great detail, leading to the development of a variety of tools that can be applied to the life sciences (Komor, Badran and Liu, 2017; Adli, 2018; Pickar-Oliver and Gersbach, 2019). Among CRISPR-based tools; nucleases, base editors, prime editors and transposases/recombinases have been engineered to allow targeted alteration of genomic DNA sequences. Class II CRISPR-Cas systems are the most

widely used CRISPR tools, which are further divided into three types based on their Cas protein variant; II- Cas9, V-Cas12 and VI- Cas13. Cas variants II and V possess RNA-guided DNA endonuclease activity and are used in genome editing tools, whilst type VI have preferential RNA-targeting endonuclease activity (Anzalone, Koblan and Liu, 2020).

Cas9 genome editing tools utilise a single guide RNA (sgRNA) engineered by the fusing of a CRISPR RNA (crRNA) and a *trans*-activating crRNA (tracrRNA) to form a single RNA molecule, which forms a ribonuclease protein complex with the Cas9 effector nuclease. The PAM sequence for Cas9 effectors is located at the 3-prime end of the protospacer on the DNA strand not complimentary to the guide RNA. The Cas9 nuclease creates a blunt double stranded break (DSB) 3bp upstream of the PAM sequence within the protospacer sequence (Garneau *et al.*, 2010). Whilst initial reports of genome editing were carried out from *Streptococcus pyogenes* derived Cas9 (spCas9) many more variants from orthologs such as *Staphylociccus aureus* and other organisms have been discovered and tested (Jinek *et al.*, 2012). These differ in their size, PAM specific sequence, editing efficiency/specificity and gRNA architecture.

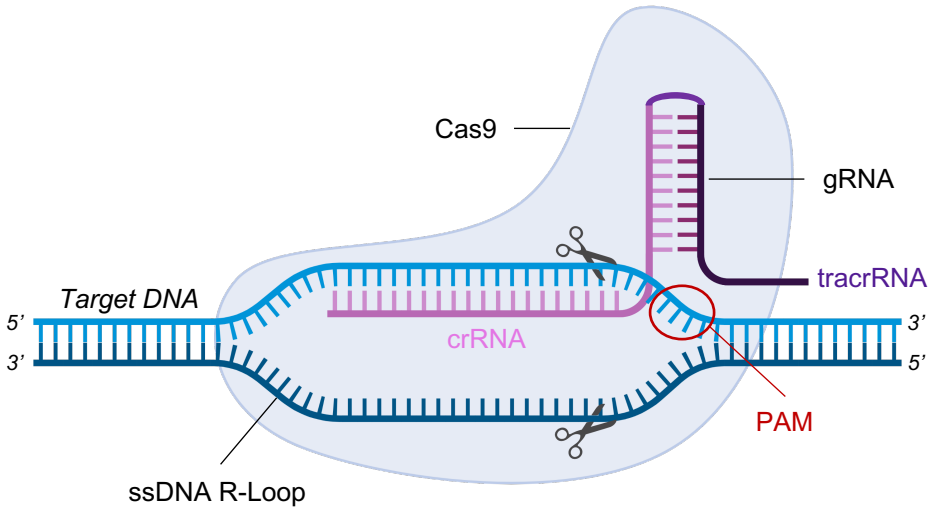


Figure 4.1 Cas9 Complex binding of target DNA: The single guide RNA (sgRNA) composed by the fusing of a CRISPR RNA (crRNA) and a *trans*-activating crRNA (tracrRNA) to form a single RNA molecule. This gRNA molecule forms a ribonuclease protein complex with the Cas9 effector nuclease which upon binding of the PAM sequence in the target DNA forms RNA:DNA heteroduplex between the sgRNA and the target DNA. Created with BioRender.com.

During the CRISPR Cas9 editing process the Cas9-sgRNA ribonuclease complex binds to the PAM sequence leading to the unwinding of the double stranded DNA and the formation of an RNA:DNA heteroduplex between the sgRNA and the target DNA. The non-target DNA strand is displaced by the gRNA and forms a single-stranded DNA 'R-loop' that is accessible to other

molecules (which is exploited in other CRISPR- tools) (Jiang *et al.*, 2016) **(Figure 4.1)**. Next Cas9 undergoes a conformational change resulting in the activation of its two nuclease domains, which are obstructed by mismatches between the target sequence and the gRNA sequence, therefore restricting the nuclease activity to DNA sequences complimentary to the gRNA sequence (Jinek *et al.*, 2014). Lastly the phosphodiester backbone of DNA is hydrolysed by the Cas9 endonuclease domains; the HNH domain which cleaves the gRNA target DNA strand and the RuvC-like-nuclease domain that cleaves the PAM containing non-target DNA strand (Jinek *et al.*, 2012). Once cleavage of the DNA strand has occurred the cell undergoes DNA repair through various pathways depending on the cell type, cell state or the nature of the DSB (Chapman, Taylor and Boulton, 2012). Two main pathways are utilised during CRISPR-Cas genome editing, the nonhomologous end joining (NHEJ) and homologous recombination (HR) (**Figure 4.2)**.

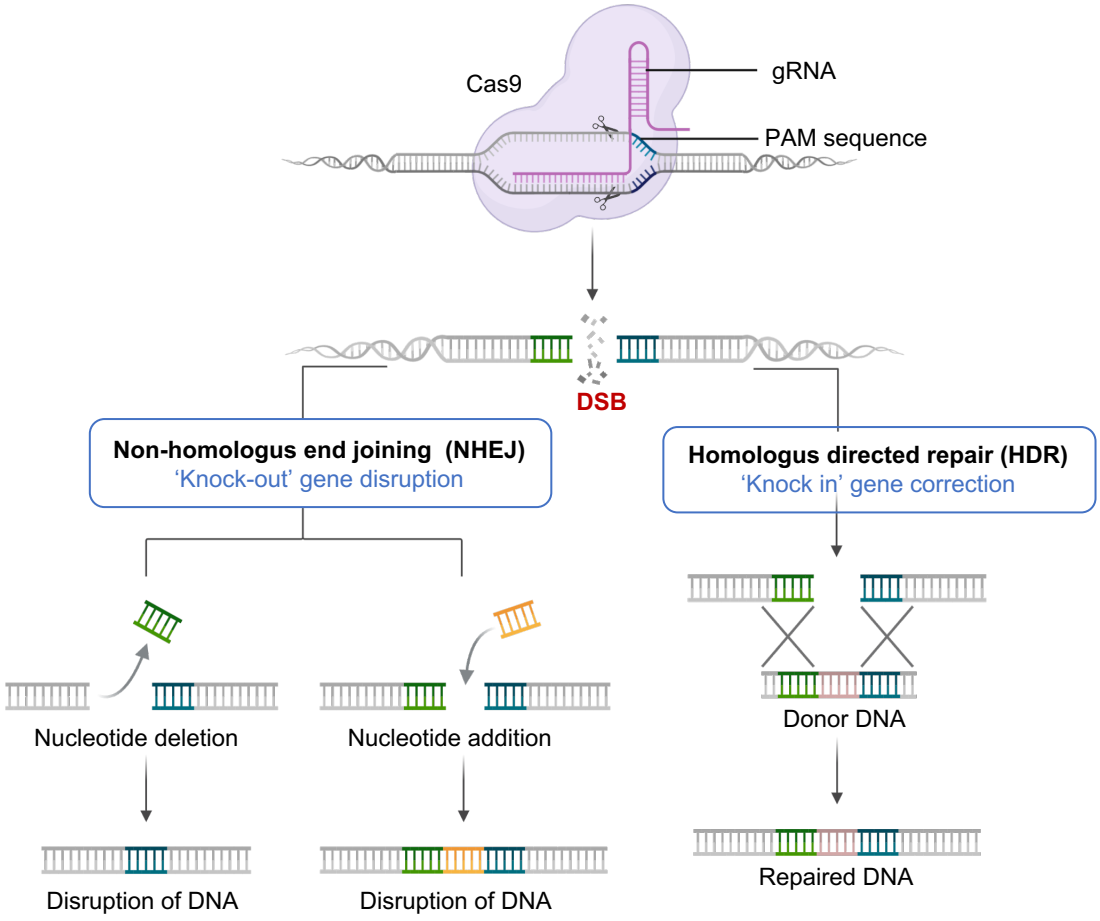


Figure 4.2 Gene editing using CRISPR/Cas9: The Cas9-sgRNA ribonuclease complex binds to the target DNA through recognition of the PAM sequence. This results in the unwinding of double stranded DNA and the formation of an RNA:DNA heteroduplex between the sgRNA and the target DNA. The non-target DNA strand is displaced by the gRNA and forms a single-stranded DNA 'R-loop' where Cas9 then induces a double stranded break. The DSB can be

repaired by the non-homologous end joining (NHEJ) repair pathway resulting in random insertions or deletions of nucleotides causing a 'knock-out' gene disruption. Alternatively, the DSB can be repaired by the Homology directed repair (HDR) pathway with the use of a donor DNA template leading to precise 'knock-in' gene editing. Created with BioRender.com.

A large proportion of experiments that utilise CRISPR-Cas nucleases aim to generate gene knockouts (KO) through disruption of target gene sequences and exploit the cells NHEJ DNA repair pathway which is most commonly activated in mammalian cells for DSB. During NHEJ the DSB are blocked from the 5' end resection and are held in close proximity by the Ku70-Ku80 heterodimer, a double stranded DNA (dsDNA) end-binding protein complex. The NHEJ pathway stimulates direct ligation of DSB ends however, the repair occurs in an error-prone manner resulting in small indels or translocations produced by the joining of different parts of the genome (Chapman, Taylor and Boulton, 2012). Though indel formation from NHEJ cannot be controlled, DNA repair profiles of edited sites have revealed that the repair is a non-random process which is determined by the protospacer sequence rather that the genomic setting (van Overbeek *et al.*, 2016). Therefore, many machine-learning models have be produced in order to predict Cas editing outcomes based on protospacer sequence (M. W. Shen *et al.*, 2018; Allen *et al.*, 2019). Nucleases have been targeted to open reading frames (ORF's), which upon indel formation result in frameshift mutations in codon sequences that abolish protein function (Shalem *et al.*, 2014; Doench *et al.*, 2016). Other experiments have used them to disrupt *cis*-regulatory elements within promoters or enhancers or to characterise non-coding RNAs (Korkmaz *et al.*, 2016; Montalbano, Canver and Sanjana, 2017). Moreover, introduction of multiple DSB within a cell, has been used to generate inversions or chromosomal translocations which occur frequently in haematological malignancies (Reimer *et al.*, 2017).

The HDR pathway is an alternative DSB repair mechanism that results in largely error free DNA repair however, it occurs far less frequently in mammalian cells than NHEJ. HDR is activated when the DSB is resected by nucleases and helicases generating 3' single stranded DNA (ssDNA) overhangs to which RAD51 recombinase assembles as a nucleoprotein filament. This structure can invade homologous duplex DNA, which is then used as a template during DNA repair (Chapman, Taylor and Boulton, 2012). HDR can be used to incorporate a variety of different genomic edits including precise insertions and deletions, point mutations and integrations of large DNA fragments. This requires a designed HDR donor template containing the specific sequence to be incorporated into the DNA.

Base editors are another CRISPR-Cas tool that can be used to convert a single nucleotide to another within the genome without the induction of a DSB or the requirement of a donor DNA

template. These editors are composed of a catalytically impaired Cas nuclease that is unable to create DSBs in the DNA which is fused to a ssDNA deaminase enzyme and in some cases contains proteins that manipulate the repair machinery. Two types of base editors have been engineered; cytosine base editors (CBE) that catalyse the conversion of C:G to T:A base pairs and adenine base editors (ABE) that catalyse A:T to G:C (Komor *et al.*, 2016; Gaudelli *et al.*, 2017). During the process of genome editing the BE Cas9-sgRNA ribonuclease complex forms the ssDNA R-loop as previously described. Conversely, no DSB is generated, and the R-loop allows the deaminase enzyme access to the target DNA sequence. The PAM-proximal nucleotides are obstructed by the binding of the Cas effector domain (approximately six nucleotides adjacent to the PAM sequence) however, the PAM-distal nucleotides are accessible to the deaminase domain (Anzalone, Koblan and Liu, 2020). At this point CBEs use cytidine deaminases to convert cytosines to uracil's which are read by polymerases as thymine or alternatively, ABE utilise laboratory-evolved TadA deoxyadenosine deaminases to convert adenosines to inosines which are read as guanines by polymerases (Komor *et al.*, 2016; Gaudelli *et al.*, 2017). As approximately 30% of known pathogenic single nucleotide polymorphisms (SNPs) are caused by C>T, A>G, T>C or G>A conversions, BE have become an invaluable tool for modelling or correcting pathogenic SNP mutations **(Figure 4.3)** (Gaudelli *et al.*, 2017; Rees and Liu, 2018; Anzalone, Koblan and Liu, 2020).

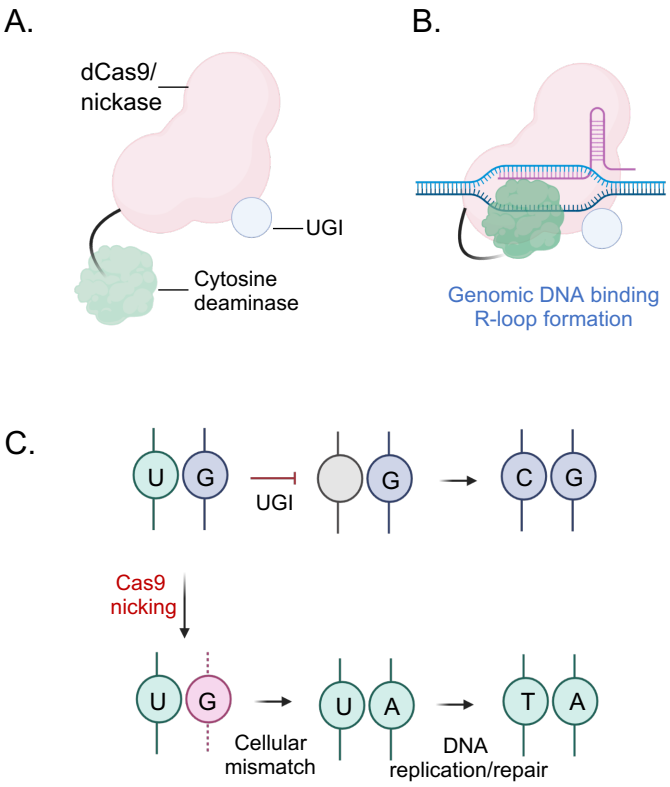


Figure 4.3: Third- generation base editor (BE3) facilitates specific guide RNA- programmed C.G >T.A modification. (A) Components of base editing workflow: CAS9 (D10A) nickase, gRNA containing locus specific sequence, Uracil DNA glycosylase (UGI) (blue) and cytidine deaminase enzyme (green). **(B)** gRNA is complexed with Cas9 and directs it to the target C in the genomic locus of interest. A ssDNA-specific cytidine deaminase enzyme catalytically deaminates cytidine nucleobases within the single-stranded area of the R-loop to uracil bases. C) Uracil DNA glycosylase (UNG) can excise uracil from the U.G mismatch, resulting in conversion back to a C.G pair. The UGI component of the base editing suppresses uracil excision. BE3 incorporates a Cas9 nickase to selectively nick the G-containing strand opposite the converted uracil. The nick guides the mismatch repair machinery to replace the G-containing strand resulted in desired T.A base pair.

Table 4.1: Base editors used to correct or model mutations

BE	Disease	Application	Cell type	Reference
CBE	Alzheimer disease	*APOE4*	Murine astrocytes	(Komor *et al.*, 2016)
	Breast cancer	Correction P53 mutation Y163C		
	Congenital disorder: Glycosylation type 1f	Correction *MPDU1* mutation L119P	Patient derived fibroblasts	(Koblan *et al.*, 2018)
		Insertion of mutation in Psen1	Murine *in vivo* model	(Sasaguri *et al.*, 2018)
	Human glioblastoma (GBM)	Insertion of *IDH1* R132H mutation	Human astroglial cells	(Wei *et al.*, 2018)
ABE	hereditary haemochromatosis	Correction Mutation C282Y Insertion of mutation increasing fetal haemoglobin (*HBG*)	Patient derived LCL	(Gaudelli *et al.*, 2017)
	B-thalassemia	Correction of mutant *HBB* allele	Engineered HEK293T cell line Patient-derived fibroblasts	(Liang *et al.*, 2017; Gehrke *et al.*, 2018)
	Androgen insensitivity syndrome (AIS)	Induction of *Ar* mutation	Murine *in vivo* model	(Z. Liu *et al.*, 2018)

Lymphoblastoid cell line (LCL)

There has been much research over the years investigating the loss of GATA2 in a variety of mouse and human models. Whilst some of these have been able to recapitulate aspects of the disease observed in human GATA2 immunodeficient patients, they do not take into account the allelic variation observed in the clinic. To date there have been over 350 genomic variations of GATA2 gene which result in many different phenotypes. Some mutations cause a quantitative GATA2 deficit whilst others lead to the production of mutant GATA2 proteins with altered expression not represented by current studies (Shimizu and Yamamoto, 2020). Pathogenesis of mutations have also been observed to differ depending on mutation location. Mutations found in the NF domain have been documented as secondary CEBPA mutated AML and activating mutations causing CML, whilst those in the CF domain cause germline initiating mutations. Each mutation has the potential to have very distinct effects on the genetic networks which control the function of stem and progenitors of the haematopoietic system.

Studies which investigate primary haematopoietic stem and progenitor cells from GATA2 haploinsufficient patients would be the ideal way to examine the molecular mechanisms underlying transformation to haematological malignancies. However, this proves difficult as these patients generally present with bone marrow hypercellularity prior to the onset of MDS or AML (Ganapathi *et al.*, 2015). Moreover, cells that can be sampled from primary MDS patients have poor growth during *in vitro* culture making them extremely difficult to work with in the lab. However, with the advent of iPS cell technology (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007), reprogramming of patient derived iPS cells has been achieved and

through initiation of differentiation programs, throughput genetic investigation of individual patient phenotypes have been achieved for many inherited bone marrow failure syndromes (BMFS) such as Fanconi anaemia (Yung *et al.*, 2013)(Liu *et al.*, 2014), Shwachman-Dimond syndrome (Tulpule *et al.*, 2013) and Familial platelet syndrome (Connelly *et al.*, 2014; Sakurai *et al.*, 2014; Antony-Debré *et al.*, 2015). However, reprogramming resistance remains a large hurdle with various mutated pathways being critical for the reprogramming stages of somatic cells. Many transcription factors, signalling pathways and microRNAs have been observed to block somatic cell reprogramming as they play key roles in either inducing or maintaining pluripotency in mouse and human cells. Other determinants of effective reprogramming are cell senescence and epigenetic state (Haridhasapavalan *et al.*, 2020). This has been observed in the reprogramming of GATA2 patient samples which were unable to capture cells carrying secondary mutations *STAG2* genetic lesion monosomy 7 (Kotini *et al.*, 2017; Jung *et al.*, 2018). Similarly reprogramming of cells from Fanconi anaemia patients is largely impeded due to the defects in the FA DNA repair pathway which is activated upon initiation of reprogramming therefore leading to decreased efficiency (Müller *et al.*, 2012; Navarro *et al.*, 2017). Another challenge of reprogramming patient samples is the heterogeneity of resulting clones carrying multiple mutations or cytogenetics. To over-come this issue specific mutations can be engineered into wild type iPSCs using CRISPR/Cas9 editing and allows specific stages of disease to be modelled with additional mutations in order to track progression of disease.

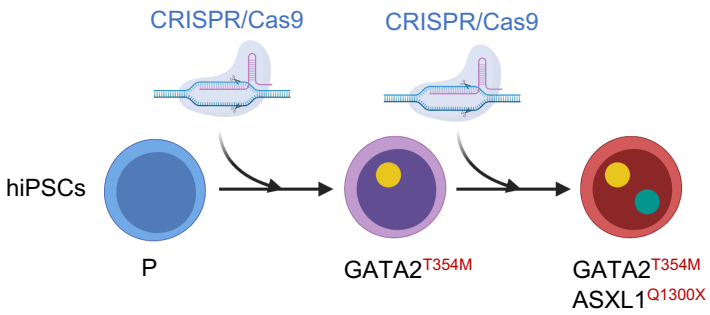
4.2 Aims and objectives

The specific aims of this chapter are:

1. Generate heterozygous GATA2 T354M mutation in human iPS cell lines using CRISPR-Cas9 nucleases.

2. Generate heterozygous GATA2 mutations T354M and R361C in individual human iPS cell lines using CRISPR-Cas9 Base editors.

3. Generate heterozygous secondary ASXL1 mutations Q1300X and R1415X observed in GATA2 patients using CRISPR-Cas9 Base editors.



Proposed CRISPR/Cas9 editing strategy: GATA2 Mutations engineered into parental human iPSC lines are used to model GATA2 immunodeficiency syndromes. GATA2 edited lines engineered to carry a secondary mutation in the ASXL1 gene models the progression towards myeloid malignancy.

4.3 Results

4.3.1 Generating heterozygous GATA2 T354M mutation using CRISPR-Cas9 nucleases

4.3.1.1 gRNA and HDR design for GATA2 patient specific mutations

To study the function of patient specific GATA2 mutations during haematopoietic development, we aimed to engineer a human iPS cell line carrying the GATA2 pThr354Met mutation using CRISPR/Cas9 gene editing. The pThr354Met GATA2 mutation (herein referred to as T354M) is one of the most prevelant mutations described throughout the literature (Bigley *et al.*, 2011; Dickinson *et al.*, 2014; Spinner *et al.*, 2014). This mutation is located in the first of five consecutive Threonine residues within the highly conserved ZF2 domain involved in DNA binding and heterodimerisation (Hahn *et al.*, 2011). This is a missense mutation where a C>T change at position c.1061 results in the amino acid change from Threonine to Methionine. To design this mutation the human *GATA2* sequence NM_032638.4 (available in the National Centre for Biotechnology Information (NCBI) GeneBank database) was used to identify the position of the single base mutation that results in the codon change p.Thr354Met observed in patients (c1061). A small section of the *GATA2* sequence (50bp upstream and downstream of base c1061) was used as the input sequence in the online tool 'crisp http://crispor.tefor.net/', which is able to predict off target scores and efficiency scores for gRNA according to specific protospacer adjacent Motif (PAM) (Haeussler *et al.*, 2016). One gRNA sequence was designed to target the T354 region of the *GATA2* gene at exon 5, on the forward strand with the PAM motif TGG, placing the c1061 base at position 14 from the predicted cleavage site, three bp from the PAM **(Figure 4.4 A).** Next, a DNA repair template was designed with the target C>T base change at c.1061 also including a CRISPR/Cas9 blocking mutation A>T at a.1077, upstream of the PAM site and left and right homology arms to the GATA2 sequence **(Figure 4.4 B)**.

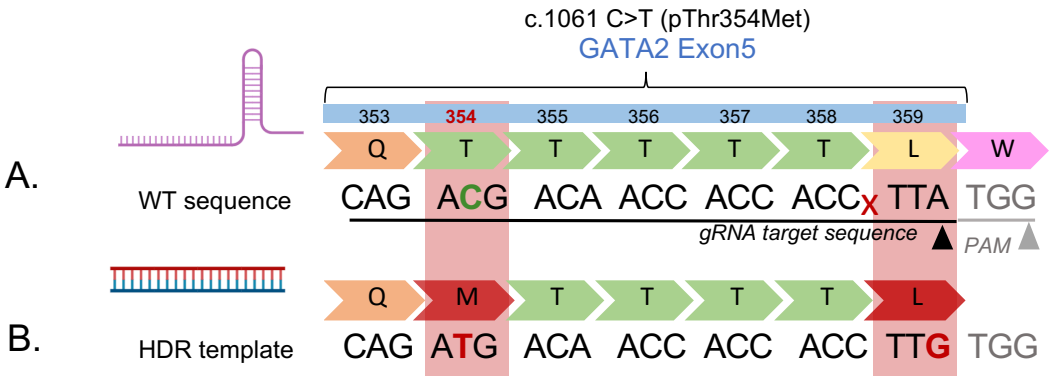


Figure 4.4: Generating human iPSC lines with GATA2 T354M heterozygous mutation using CRISPR/Cas9 genome editing. (A) Strategy of GATA2 T354M mutation; synthetic CRISPR RNA (cRNA) was generated to target the T354 region located in C-finger domain, at exon 5 of the human GATA2 gene, nucleotides in grey are the protospacer adjacent motif (PAM) 'TGG'. A 130bp HDR template was designed to incorporate the C>T nucleotide change at c.106, resulting in the 354 codon change from threonine to methionine. A CRISPR/Cas9- blocking mutation was incorporated upstream of the PAM site at a.1077 resulting in the same codon.

4.3.1.2 Delivery of Cas9-gRNA ribonuclease complex via Electroporation

Delivering the CRISPR-Cas9 components as a Cas9-gRNA ribonuclease complex was chosen as it allows for rapid cleavage of target DNA, by-passing transcription and translation steps required in plasmid or mRNA delivery methods (Liang *et al.*, 2015; DeWitt, Corn and Carroll, 2017). Furthermore, rapid clearance of the Cas9 RNP occurs via protein degradation pathways shortly after transfection (~72hrs), decreasing the time that Cas9 is available for off-target cleavage (Kim *et al.*, 2014; Vakulskas and Behlke, 2019). The gRNA sequence was synthesised as TrueGuide™ Synthetic crRNA by Thermo Scientific and annealed to the TrueGuide™ a custom Synthetic trRNA (which mediates the interaction with CAS9) to form a gRNA:tracrRNA duplex. The DNA repair template was synthesised as a two complementary long oligonucleotide strands (Sigma) and annealed together to form a double stranded oligonucleotide HDR template.

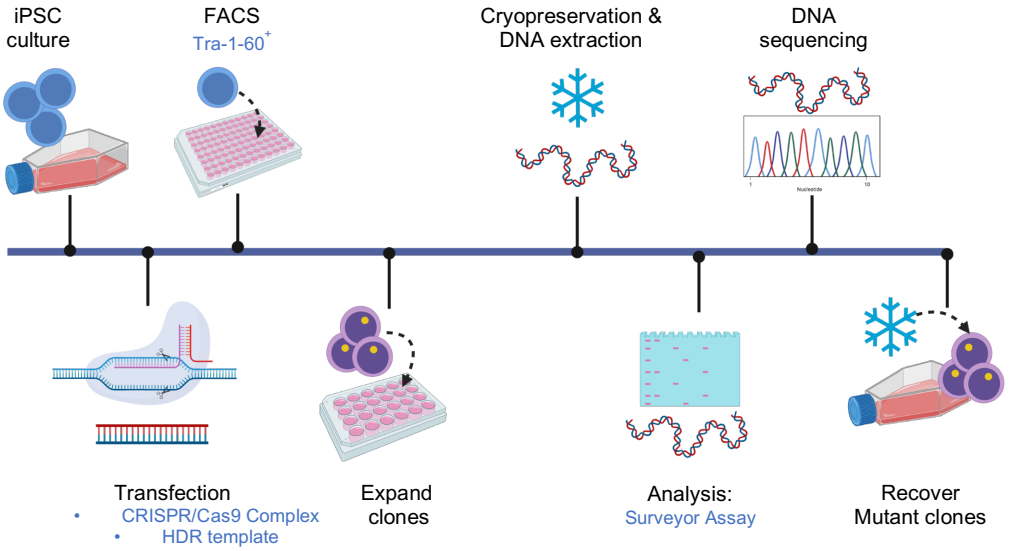


Figure 4.5: Schematic illustration of experimental plan. Human iPS cells were transfected with the CRISPR/Cas9 complex (True guide [T354M]:TrueCut Cas9 Protein v2) by electroporation, 72 hours later cells were single cell sorted based on TRA-1-60 expression, expanded, split for cryopreservation or alternatively for genotyping using the surveyor assay and confirmed with next generation sequencing. *Image created using BioRender platform.*

Human iPS cells were cultured and the Cas9-gRNA complex delivered into the cells by nucleofection, an electroporation-based transfection method which has been shown to result in high transfection efficiencies in human PSC's over other methods. After the Cas9 protein had degraded (72hrs post transfection) the cells were sorted for their expression of pluripotent marker TRA-1-60, as individual cells per 1 well of a 96 well plate. The gating for flow sorting was set to include only very highly expressing TRA-1-60 transfected iPS cells to ensure a truly pluripotent population was isolated. Individual clones appeared 12-14days post sorting which maintained hPSC-like morphology and a total of 104 clones were manually expanded and split into two aliquots; one for cryopreservation of the clone and the second for genotyping of the GATA2 T354 region **(Figure 4.5)**.

4.3.1.3 Screening of clones for T354M mutation

To detect the recombination event leading to the HDR template insertion into the cells, the Surveyor assay was used to scan clones for polymorphisms in heteroduplex DNA. The assay employs the Surveyor™ endonuclease which cleaves mismatches in heteroduplex DNA due to the presence of single nucleotide polymorphisms (SNPs) or small indels, cleaving both DNA strands downstream of the mismatch (Qiu *et al.*, 2004; Vouillot, Thélie and Pollet, 2015). The

GATA2 gene was amplified by PCR in individual clones using primers 500bp up and downstream of the edit site c.1061, digested and band pattern assessed by gel electrophoresis. The activity of the endonuclease showed 53% (56/107) of clones with un-cleaved amplicons observed as a band at 938bp, along with two cleaved products at 477bp and 461bp corresponding to heteroduplex DNA at theT354 location. Three clones were observed to have two un-cleaved bands of ~930bp and 800bp with cleaved bands either at ~500bp or alternative amplicon sizes suggest a mixed clone resulting from more than one cell being sorted during clone propagation. Furthermore, from the total clones analysed only two produced a single un-cleaved amplicon at 938bp, suggesting a wild-type genotype at both alleles and a 96% (102/107) Cas9 cutting efficiency at the gRNA target sequence **(Figure 4.6 A)**.

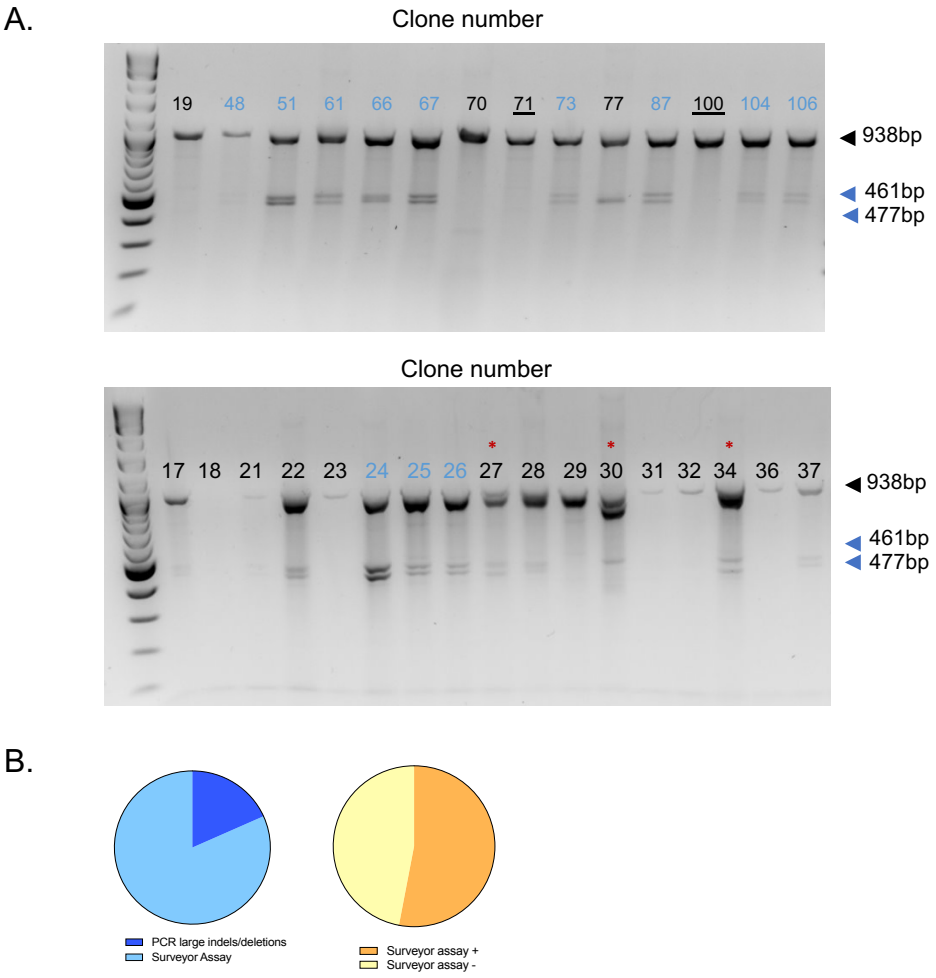


Figure 4.6: Screening CRISPR/Cas9 clones using the Surveyor cleavage assay. (A) Surveyor mismatch cleavage assay was carried out on recovered clones, 500bp up and down stream of T354 edit site was amplified and digested with the surveyor endonuclease to detect mismatched DNA. Products were resolved by electrophoresis on a 1% [w/v] agarose gel in 1x TBE buffer using a 2LogDNA ladder to detect molecular weight. Activity of the endonuclease causing cleavage at heterozygous DNA at the T354 region of the GATA2 gene resulted in an un-cleaved amplicon (938bp) and two cleaved products at 477bp and 461bp (blue), wild-type clones result in one un-cleaved amplicon (938bp). Clones with mixed populations of cells are observed having more than one amplicon of mixed size and alternative banding pattern **(*). (B)** Summary of the efficiency of Cas9 editing at the T354 site of the GATA2 determined by the surveyor assay from 104 clones screened

Further analysis of nucleotide sequence was carried out for clones presenting with potential heterozygous mutations at the T354 location from the Surveyor assay. Genomic DNA was sent for next generation sequencing and analysed for incorporation of the HDR template. DNA sequence analysis revealed seven clones had successfully undergone HDR of the DSB on one of the GATA2 alleles, observed by the incorporation of the C>T change at c.1061 and the

A>T CRISPR/Cas9 blocking mutation at a.1061. Unfortunately, both of these clones had undergone DSB repair through the NHEJ pathway resulting in large insertions and deletions to the second allele causing disruption to the gene **(Figure 4.7 A)**. The remaining clones analysed, had large unwanted deletions or insertions on both GATA2 alleles indicative of DSBs repaired by the NHEJ pathway **(Figure 4.7 B)**.

Thus, from this initial experiment we can conclude, although delivering the Cas9-gRNA complex via electroporation produced extremely high CAS9 editing at the T345 GATA2 locus in human iPS cells, successful editing of our specific nucleotide changes were restricted by the efficiency of the cell to repair the DSB through the NHEJ repair pathway over the HDR pathway.

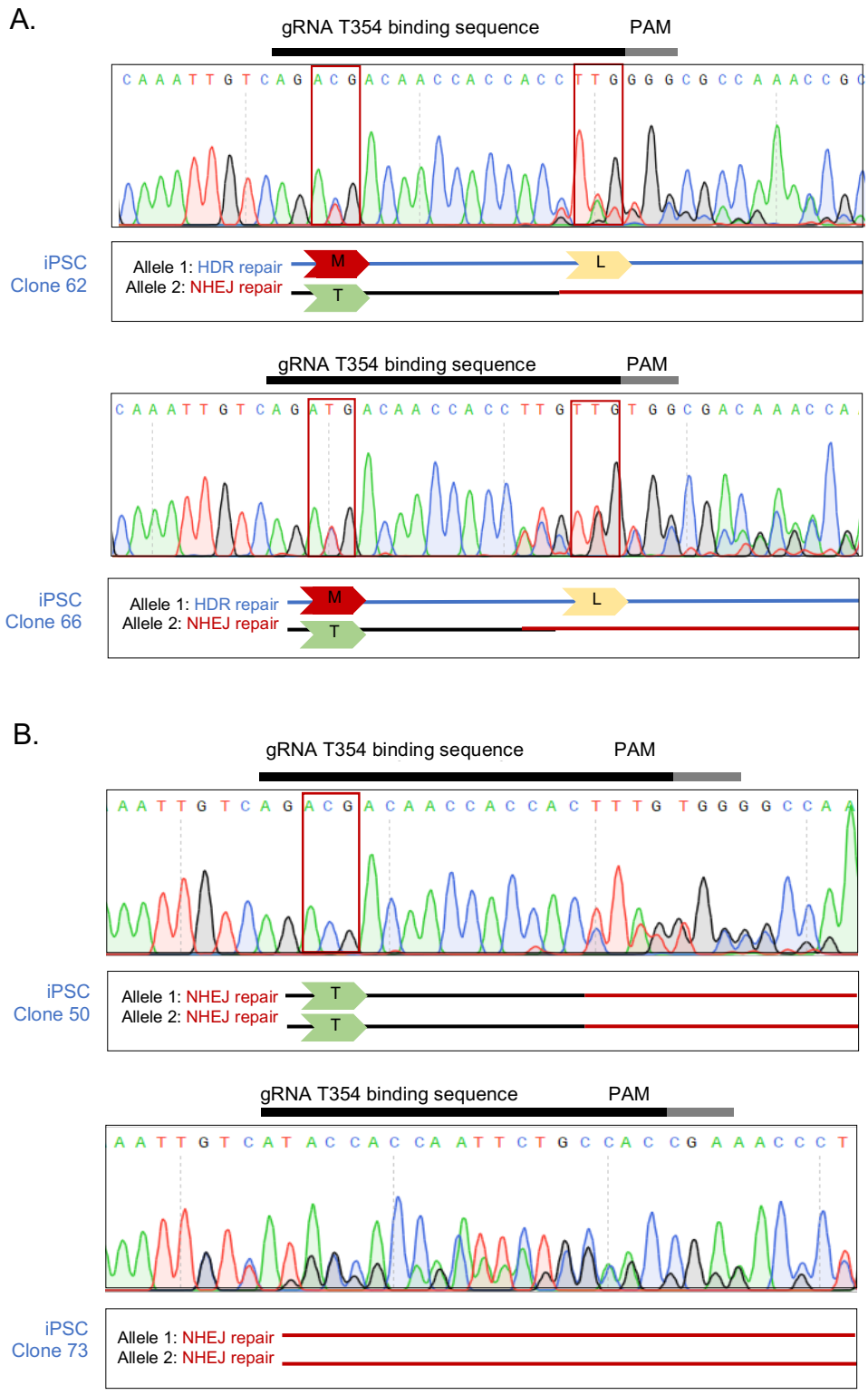


Figure 4.7: Mutations incorporated at the T354 site in human iPS cells using CRISPR/Cas9 editing. Representative next generation sequencing chromatograms of clones positive in the Surveyor assay. **(A)** Chromatogram analysis showed clones 62 & 66 to have a C>T change at c.1061 and the A>T change at a.1077 on one allele. However, the second allele in each clone had been disrupted by insertions/ deletions at the double strand break site. **(B)** Clones 73 and 50 represent the majority of cloned screened, with disruptions of the gene on both alleles from large insertions and deletions.

4.3.2 Generate heterozygous GATA2 T354M and GATA2 R361C mutations using CRISPR-Cas9 Base editors.

Due to the lack of successful editing from our previous experiment, we sought to employ an alternative genome editing technique that would introduce heterozygous point mutations in a more efficient manner. We implemented the state-of-the-art and highly efficient "single base editing" technique newly developed by Komer *et al.* which utilises a third generation of base editor (BE3) composed of dCAS9, tethered to a cytidine deaminase (APOBEC1) to facilitate a C to T conversion in the selected guide RNA region and an uracil glycosylase inhibitor (UGI) **(Figure 4.3)**. Furthermore, as there is no double strand break created by the Cas9 enzyme, the NHEJ DNA repair machinery is bypassed and stochastic insertions and deletions in genomic region of interest are greatly reduced.

4.3.2.1 gRNA design for GATA2 patient specific mutations

These base editors also have limitations as certain design criteria must be met for successful editing. As cytosine deaminase facilitates the conversion of cytosines to thymidine's, only C>T or A>G (designing a guide on the antisense strand) SNP change can be modelled with this technique. Moreover, efficient editing by APOBEC requires the presence of the PAM motif NGG, which also places the target nucleotide at position 4-8 of the PAM-distal end of the protospacer- the "editing window". This requirement limits the locations in the human genome that can be efficiently targeted by BE3 due to the limiting number of PAM sequences. Additionally, due to the high activity of APOBEC, all cytosines located in the five- nucleotide window will be changed to uridines (though with varying efficiencies) (Kim *et al.*, 2017; Rees and Liu, 2018; Anzalone, Koblan and Liu, 2020). Nevertheless, improved BE3 base editors have been engineered containing mutated forms of cytidine deaminase domains, causing narrowing of the editing window to 1-2 nucleotides in length such as; YE1-BE3, YE2-BE3 and YEE-BE3. These new editors prevent neighbouring C's from also being targeted along with the intended SNP change, whilst also doubling the number of disease-associated Cs which can be modelled using this technique (Anzalone, Koblan and Liu, 2020). Taking these limitations into consideration, all known familial GATA2 mutations were screened to assess for their suitability, according to the aforementioned design parameters, using the Benchling CRISPR Guide RNA Design platform (ENSG00000179348 GATA2 reference). Guide availability was restricted to the modelling of two GATA2 mutations T354M and R361C.

The guide designed to target the T354 region (as previously described) when used with the base editor BE3-GAM, positioned the editing window at bases 4-8, resulting in two edits at positions C_4 (target cytosine) and C_7 (bystander edit) which would result in a second codon change from Thymidine to Isoleucine **(Figure 4.8 A)**.

The second mutation chosen to model was the *p.Arg361Cys GATA2* mutation (herein referred to as R361C) reported as another of the most prevelant mutations described throughout the literature (Bigley *et al.*, 2011; Dickinson *et al.*, 2014; Spinner *et al.*, 2014). This is a missense mutation located in the ZF2 domain where a C>T change at position c.1081 results in the amino acid change from Arginine to Cysteine **(Figure 4.8 A)**. Two guide RNA were designed; the first to be used in combination with the base editor BE3-GAM which would result in one edit at C_7 (target cytosine). The second guide was designed one base upstream of the previous gRNA and when used with the modified base editor YE1-BE3 would provide a reduced editing window at bases 5-7, placing the target cytosine in the centre with no bystander edits **(Figure 4.8 B).**

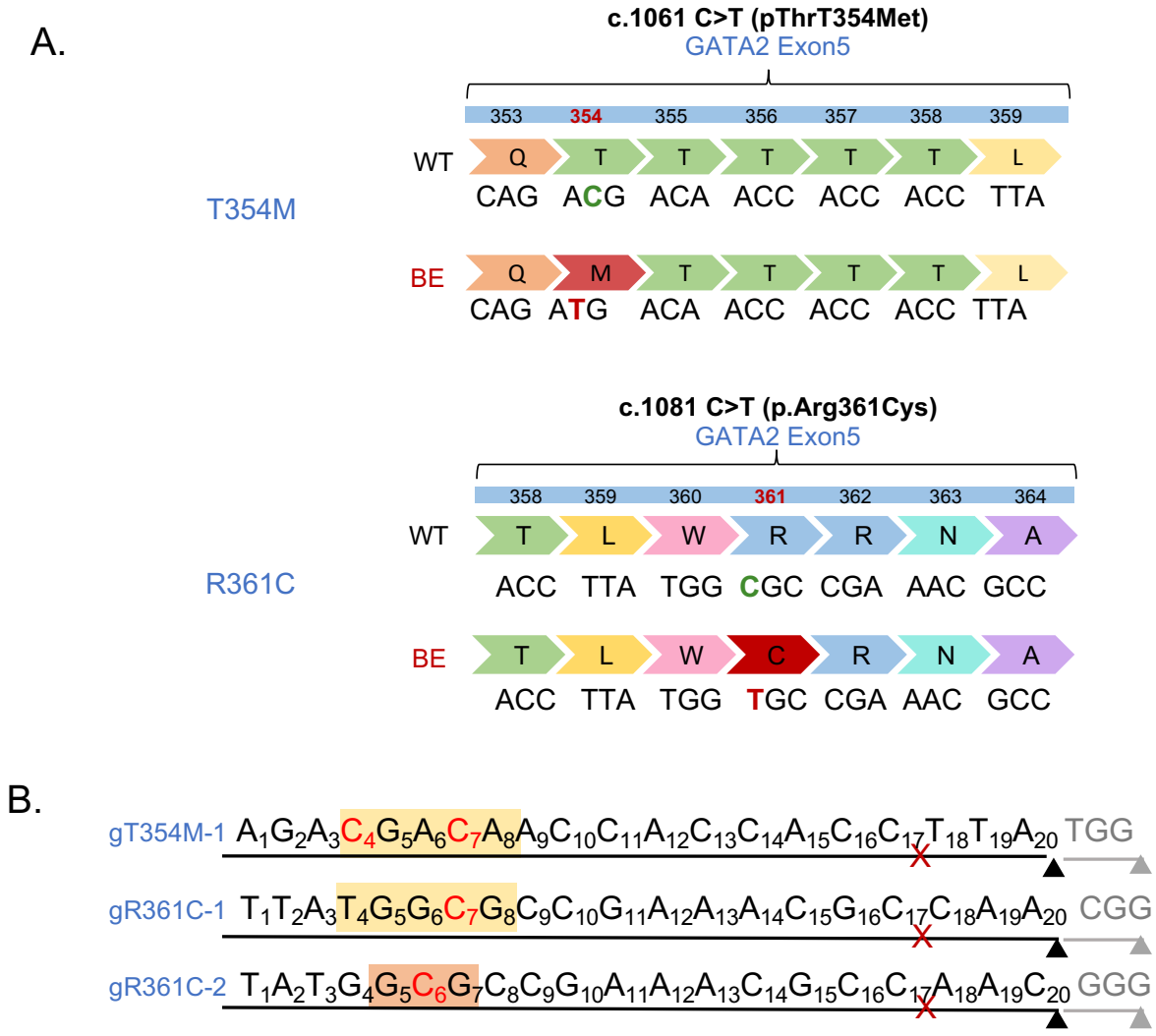


Figure 4.8: Base editing design of GATA2 mutations T354M and R361C. (A) Location of GATA2 mutations; *T354M* as previously described, *R361C* is located in C-finger domain, at exon 5 of the human GATA2 gene and results from a C>T change at c.1081 leading to the codon change from Arginine to Cysteine. **(B)** gRNA sequences designed for GATA2 mutations *T354M* and *R361C* shown in black, with the protospacer adjacent motif (PAM) in grey. Cytosine base editor activity windows are represented in yellow for unmodified BE3 editor and orange for modified YE1-BE3 editor and all possible cytosines that can be deaminated to thymine's are shown in red. The location of the nick in target DNA backbone is indicated by the red x, base positions are numbered relative to their PAM-distal end of the gRNA.

4.3.2.2 gRNA design for ASXL1 secondary mutations

To investigate the progression of GATA2 immunodeficiency, we aimed to model the acquisition of secondary ASXL1 mutations which correlate with myeloid transformation. These

ASXL1 mutations are mostly heterozygous, frame shift or missense mutations that result in premature termination codons, located predominantly in exon 13 (West *et al.*, 2014; Alvarez Argote and Dasanu, 2018). A gRNA sequences was designed to target p.Q1300X C>T mutation at c.3898, resulting in the codon change from Glutamine to stop codon TAA **(Figure 4.9 A)**. When used with BE3-GAM editor would produce an edit at C_4 (target cytosine) with no bystander edits. The second gRNA was designed to targeted R1415X, C>T mutation at c.4243 causing the codon change from Arginine to a stop codon TGA. When used with BE3-GAM it places $C_{4\text{-}7}$ in the editing window causing two edits change from Proline to TTT at P1414 and the codon change from Arginine to stop codon TGA **(Figure 4.9 B)**. Following the design of the guide sequences complementary oligosaccharide GATA2 or ASXL1 gRNA sequences were annealed together and cloned into the U6gRNA vector using BbsI digestion sites **(section 2.3, Figure 2.4)**.

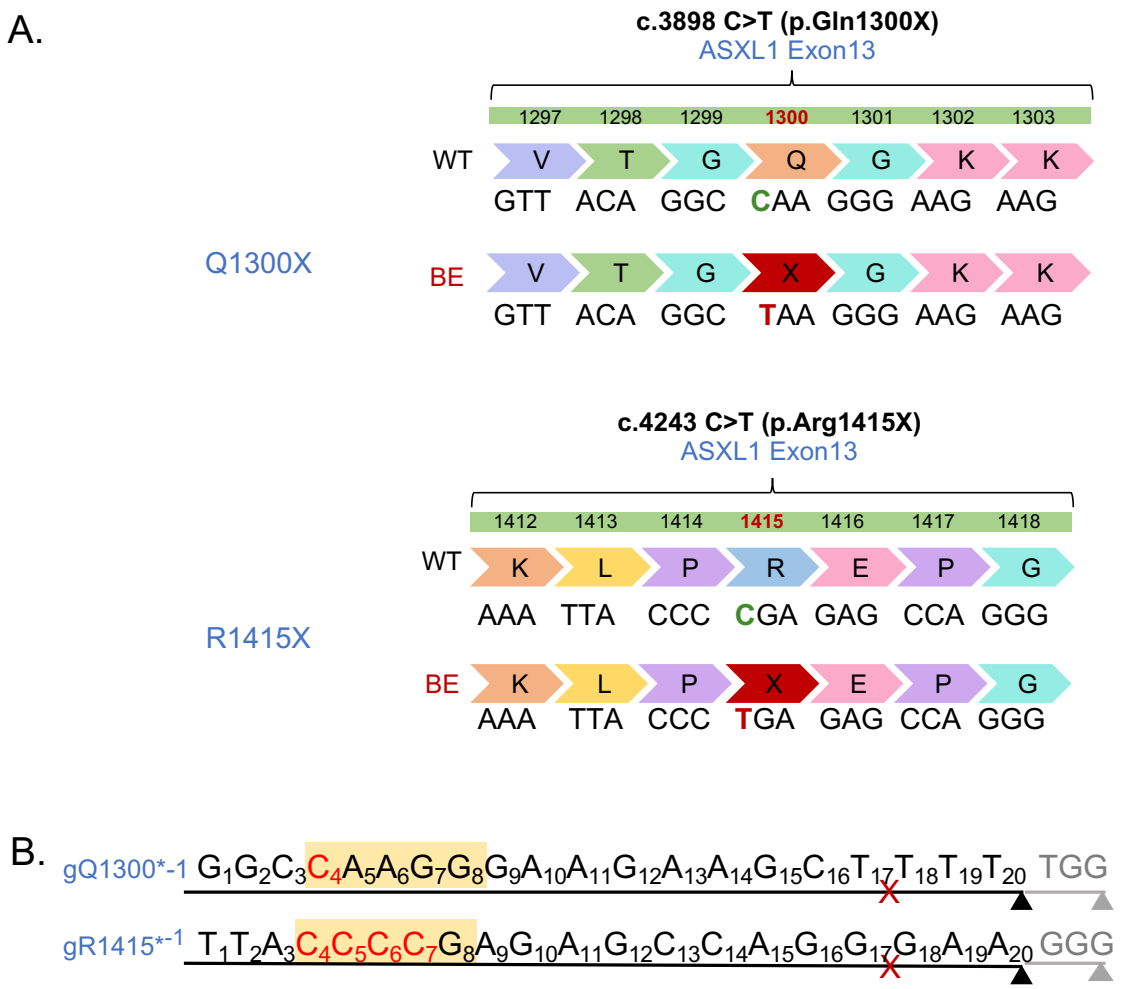


Figure 4.9: Base editing design of secondary mutations in ASXL1; G1300X and R1415X. (A) ASXL1 mutations are located in exon 13, *Q1300X* is a C>T change at c.3898 resulting in the codon change from Glutamine to a termination codon TAA, *R1415X* results from a C>T change at c.4234 resulting in the codon change from Arginine to a termination codon TGA. **(B)** gRNA sequences designed for ASXL1 mutations *Q1300X* and *R1415X* are shown in black with the protospacer adjacent motif (PAM) in grey. Cytosine base editor activity windows are represented in yellow for unmodified BE3 editor and orange for modified YE1-BE3 editor and all possible cytosines that can be deaminated to thymine's are shown in red. The location of the nick in target DNA backbone is indicated by the red x, base positions are numbered relative to their PAM-distal end of the gRNA.

4.3.2.3 Delivery of T354 and R361 gRNA's and Base editors for editing at the GATA2 genomic loci.

To incorporate our specific edits, plasmids containing either the cloned T354M or R361C-G2 gRNA carrying a G418 antibiotic resistant marker, and the base editors BE3-GAM and YE1-BE3 were delivered via electroporation into the human EPI-iPSC line. 72hrs post transfection cells were selected for gRNA expression with G418 antibiotic for a further five days followed

by single cell sorting into single 96 wells. These clones were expanded into cultures for cryopreservation and cultures for genotyping from extracted genomic DNA **(Figure 4.10)**. The Taqman SNP genotyping assay was performed on 50 clones recovered from base editing experiments using the gT354M and BE3-GAM and 60 clones recovered from gR361C-2 and YE1-BE3. All samples were genotyped alongside annealed oligonucleotide sequences containing either; an unmodified T354/R361 sequence (homozygous control), sequences containing the specific SNP change (homozygous mutant) or a 1:1 mixture of the two oligonucleotides (heterozygous mutant) as reference samples. We found that all analysed clones showed amplification signals of two wild-type alleles for both genotypes analysed; $T354M^{(C/C)}$ and $R361C^{(C/C)}$ clustering with homozygous control reference samples, revealing that none of the isolated clones had successfully undergone I C>T base changes by the cytosine deaminase enzyme. This was further confirmed by DNA sequencing a selection of clones, which all revealed unmodified wild type sequences **(Figure 4.11)**.

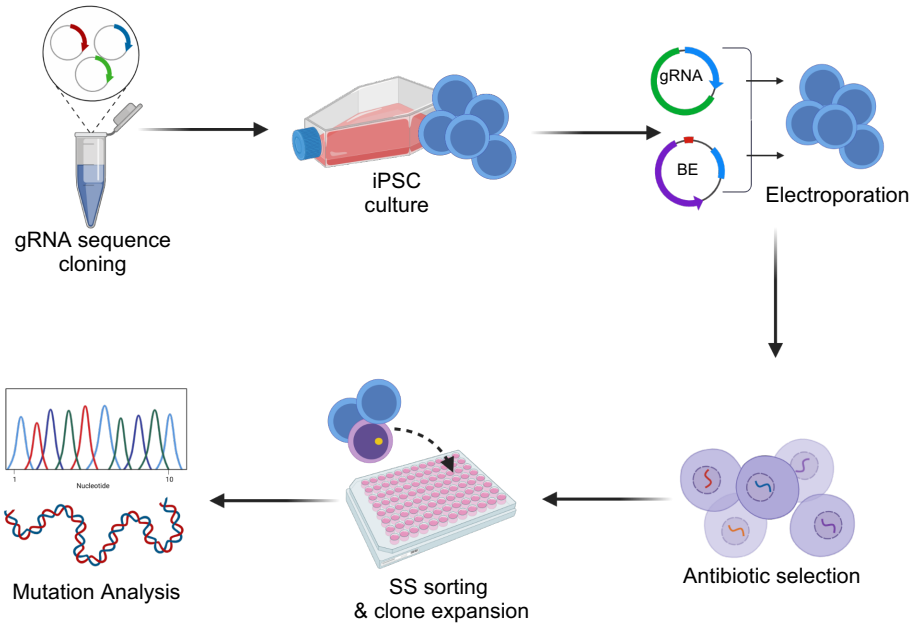


Figure 4.10: Base editing experiment workflow. gRNA sequences were cloned into the phU6-gRNA vector using BbsI restriction enzyme sites. Human iPS cells were maintained on rhlaminin coated plates in stemflex media, harvested, electroporated using the Neon system with cloned U6gRNA and base editor plasmids (phCMV BE3-Gam/ pBK-YE1-BE3 depending on target cytosine location within the editing window (figures 3.7C), 72hrs post transfection cells were selected using G418 (450 μg/mL), 5 days later cells were single cell sorted and expanded. Individual clones were genotyped for base edits resulting in desired mutations using the TaqMan SNP genotyping assay and positive clone subsequently sent for next generation sequencing. *Image created using BioRender platform.*

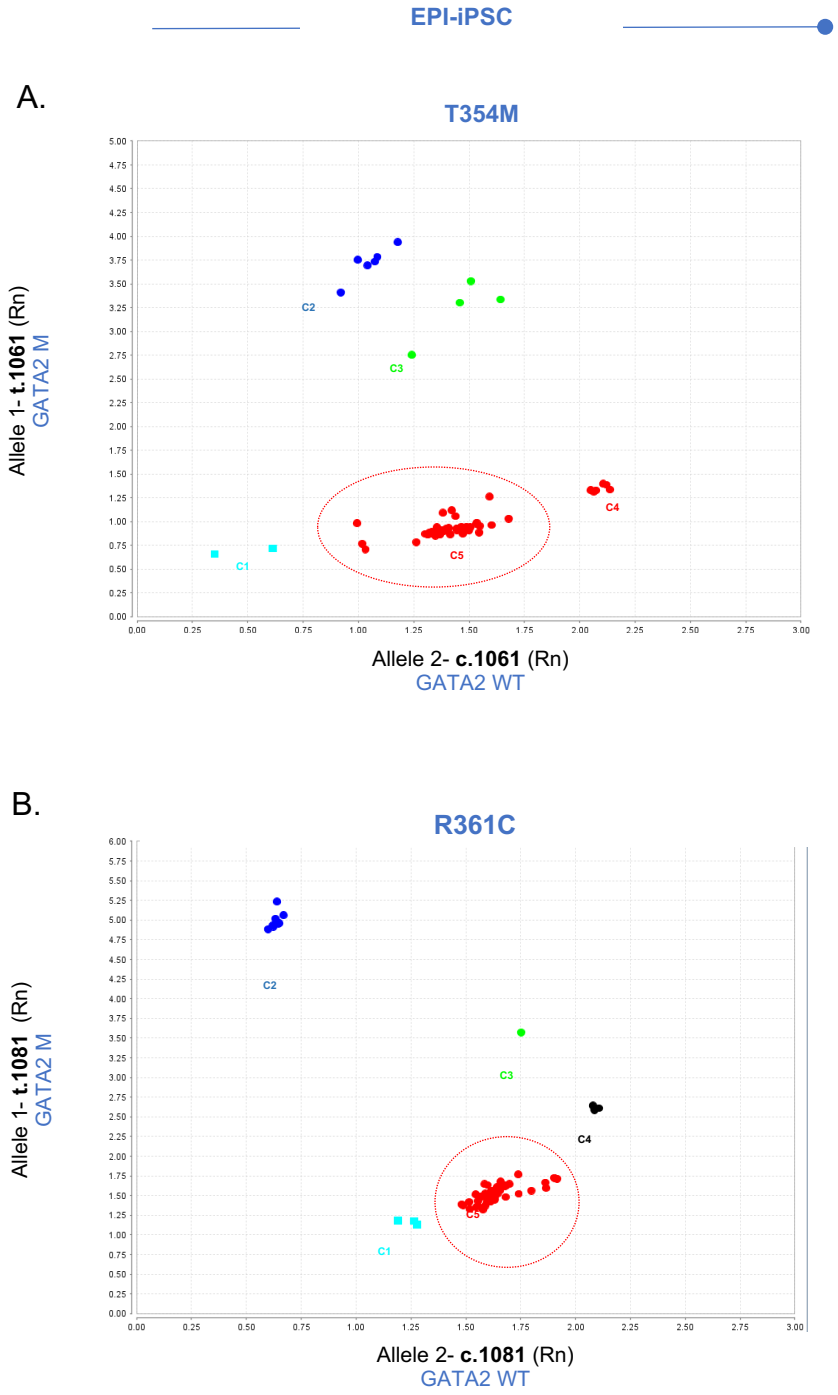


Figure 4.11: Screening of clones using the TaqMan SNP genotyping assay. The Taqman SNP genotyping assay was used to screen for GATA2 mutations **(A)** T354M$^{(-/+)}$ and **(B)** R361C$^{(-/+)}$ in the human EPI-iPSC line. All samples were genotyped alongside reactions with no DNA template (C1), annealed oligonucleotide sequences containing either; an unmodified T354/R361 sequence (homozygous control) (C4), sequences containing the specific SNP change (homozygous mutant) (C2) or a 1:1 mixture of the two oligonucleotides (heterozygous mutant) (C3) as reference samples appearing as distinct clusters. All clones analysed amplified two wildtype GATA2 alleles shown as samples (C5) clustering near C4 samples.

4.3.2.4 Optimisation of control experiments using the RNF2 genomic loci.

To assess whether the lack of editing observed at the GATA2 genomic loci in our previous experiment was due to inefficient transfection efficiency or due to limitations of the base editors to sufficiently edit our specific sequence, we modified the base editor used. Whilst we are able to select cells that are transiently expressing the gRNA plasmid using G418, we are unable to assess whether the base editor plasmids are also being expressed within the same cell as the gRNA as this plasmid does not carry a selection marker. We therefore changed our base editor to the BE4-MAX plasmid which incorporates a fourth generation base editor (BE4)-P2A-GFP fusion, enabling GFP selection which directly relates to BE4 production (Koblan *et al.*, 2018). Furthermore, as the efficiency observed from these previous base editing experiments at the GATA2 loci were surprisingly low in relation to what has been observed in the literature, we aimed to optimise the transfection using the RNF2 genomic loci, documented as a highly editable site using cytosine base editors. gRNA sequences for RNF2 outlined by David Liu, were cloned into the gRNA plasmids as previously described and delivered into the human iPSC line along with the BE4-MAX plasmid by electroporation, EPI-iPS cells were also transfected with DNA encoding the PWP1-GFP plasmid which have resulted in high transfection efficiencies using other delivery methods (Data not shown). We observed a low GFP expression in cells co-transfected with gRNA-RNF2 and BE4-MAX plasmids (2.15%) along with a control plasmid PWP1-GFP (2.49%) **(Figure 4.12)**. As these preliminary results indicated that the transfection required further optimisation, the next step was to continue with a 24well optimisation protocol to find the optimum electrical parameters for efficient delivery of our plasmids by electroporation into our cell line.

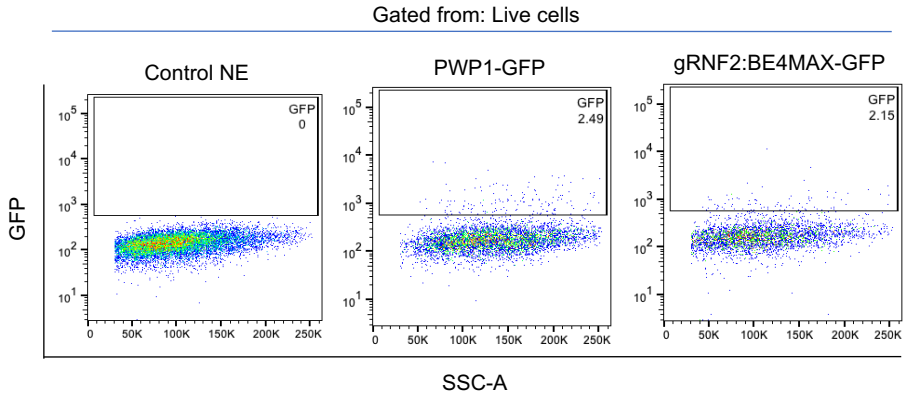


Figure 4.12: Delivery of BE4-MAX GFP plasmid by electroporation: Preliminary data from initial experiment of human EPI-iPSC line transfected with the BE4-MAX GFP plasmid. Cells were cultured to 60% confluency and harvested into a single cell suspension, $1X10^6$ cells were resuspending with plasmid DNA in electroporation resuspension buffer and transfected using recommended human ESC programme: 120V, 30s and 1 pulse. Samples were ran alongside a no electroporation control (NE) to

monitor cell viability and a sample transfected with the PWP1 control plasmid known to give high transfection efficiencies.

Unfortunately, due to the closure of the laboratory as a result of the COVID19 pandemic we were unable to continue with the base editing optimisation. During this closure period we realised that setting up the patient specific mutations in our human iPSC lines would be difficult to come to fruition in the remaining time of the PhD. Therefore, the closure period was used to plan and design two alternative approaches to investigate GATA2 disease progression outlined in **Chapter 5**.

4.4 Discussion

Disease modelling of hereditary haematological malignancies using iPS technology, has become an invaluable tool for understanding key mechanisms of disease initiation from development and its subsequent progression. In this chapter we aimed to model patient specific GATA2 mutations along with the introduction of secondary mutations in the ASXL1 gene, to investigate GATA2 immunodeficiency and its progression towards MDS/AML using iPS cells and CRISPR/Cas9 genome editing.

In our first gene editing approach we attempted to model the GATA2 T354M$^{(+/-)}$ mutation in the human EPI-iPSC line using a CRISPR/Cas9 knock-in approach. In these initial experiments we found that modifying one allele through HDR DNA repair using a DNA template containing a specific SNP, whist leaving an unmodified allele, to be very difficult as shown by the high efficiency of biallelic editing at the T354 loci. Though many groups have reported effective engineered knockout using CRISPR/Cas9 cleavage, repaired through the NHEJ pathway (Kaustio *et al.*, 2021; Khaled *et al.*, 2021), creating specific lines using the 'knock-in' approach remains limited by the efficiency of cells to undergo DNA repair through the HDR pathway (~0.1-5%) (Jasin and Rothstein, 2013). The selection of which DSB repair pathway the cell undertakes largely depends on DSB resection, a nucleolytic process that converts DSB ends into 3'-single stranded DNA overhangs (Symington, 2016). NHEJ factors such as 53BP1, promotes direct joining of DSB by protecting DNA ends from resection (Bothmer *et al.*, 2010; Gupta *et al.*, 2018). Limited resection results in exposed regions of sequence microhomology which promote repair through microhomology-mediated end joining (MMEJ) (Sfeir and Symington, 2015; Wang and Xu, 2017), whilst extensive resection generates long 3'single-stranded DNA tails required for HDR (Jasin and Rothstein, 2013; Jasin and Haber, 2016). Another important influence of HDR is cell cycle phase as many HR factors are simply not active or have ineffective expression in G1/G0 cells unlike the NHEJ pathway which dominate repair in G1, S and G2 phases (Lin *et al.*, 2014; Jasin and Haber, 2016).

In line with the literature during this experiment we report a HDR efficiency of 3.36% (7/208) in the human EPI-iPSC line. Our design followed key parameters for increasing HDR efficiency in human iPS cells, as outlined in the study lead by Paquet *et al* (2016). Our HDR template incorporated a Cas9- Blocking mutation (BM) in the codon upstream of the PAM site, preventing re-cutting by Cas9 at a successfully edited locus and its subsequent repair by the NHEJ pathway, which proved very efficient as we were able to isolate clones that had successfully incorporated our HDR template with no further indel formation as a result of

secondary binding of the gRNA and Cas9 cleavage. Additionally, the authors also described there to be an inverse relationship between mutation incorporation rate and its distance from the DSB site, which can be utilised to predict the control of zygosity. In compliance with the suggested theory that introduction of heterozygous HDR templates were more efficient when the intended mutation is placed further away from the break site, our gRNA sequence was designed with our C>T mutation at position four and the predicted Cas9 cleavage site located at position 17. The bias towards bi-allelic editing observed is likely due to the high efficiency of the Cas9 protein. To overcome this issue in future experiments we could potentially reduce the amount of Cas9 protein delivered into the cells to reduce the Cas9 cutting efficiency. This could result in a reduced editing efficacy allowing some alleles to remain unedited as heterozygous clones. Nevertheless, in reducing cas9, this would consequently result in a reduction of HDR events, significantly increasing the number of clones needed to be screened, making manual isolation and expansion unmanageable in our current laboratory.

Many groups have investigated various methods to inhibit key proteins in the NHEJ pathway to increase the HDR mediated DSB repair. Small molecules such as the putative inhibitor of ligase IV- SCR7, proteins of the adenovirus4 (Ad4); E1b55k and E4orf6 and small interfering RNA (siRNA) have been used as NHEJ block-aids (Li and Ding, 2010). However, whilst 3-7 fold increases on HDR repair have been documented, researchers have also observed this to cause substantial reduction in the expression of the target gene (M. Liu *et al.*, 2018). Other attempts have focused on blocking the cell cycle at certain phases which favour HDR such as work conducted by Lin *et al.,* 2004 which showed that treatment with Nacodazole resulted in the halting of cells in the G2/M phase, where HDR factors are expressed (restricted to S and G2 phases) and when replication is completed and sister chromatids are available for use as repair templates (Lin *et al.*, 2014). Nevertheless, these chemical manipulations can be harmful for the cells long term and can be difficult to employ (Lin *et al.*, 2014).

One possible method to improve HDR efficiency in the gene editing experiments we undertook is to use one of the above NHEJ blockades or cell cycle inhibitors along with supplying equimolar mixing of two HDR templates: one containing the mutation of interest and our Cas9-BM and a second with a wild type sequence also containing a Cas9-BM. Through decreasing NHEJ repair mechanisms, we would increase the HDR efficiency and therefor are more likely to also increase the risk of clones with biallelic HDR alleles which can use both HDR templates to achieve a GATA2 T354M $^{(+/-)}$ clone. Despite significant progress to improve HDR activation over the NHEJ pathway during genome CRISPR/Cas9 editing, current strategies still remain laborious. Based on the limitations discussed from these alternative approaches, an alternative way to create our specific heterozygous GATA2 mutations through HDR, would be

to carry out a second round of CRISPR/Cas9 editing in the isolated clones that have incorporated the HDR template on one allele **(Figure 4.4 A)**. A gRNA sequence could be designed to recognise the disrupted second GATA2 allele, along with the design of a HDR template with a wild-type GATA2 sequence which spans the length of the disrupted sequence from indel formation with sufficient homology arms. Importantly, the gRNA would not recognise the wildtype sequence preventing Cas9 from re-cutting the mutated allele and given our previous HDR efficiency (3.36%) we would be able isolate around seven clones with our desired GATA2 T354M$^{+/-}$ genotype.

During this chapter we also investigated a second approach to model these GATA2 mutations exploiting base editors developed to introduce specific C>T conversions in the genome without inducing DSBs, thus by passing the limiting factor of HDR mediated repair. As mentioned previously, base editing requires precise design parameters consisting of; editing window placement within the guide, relative to target base mutation, restriction of C>T or G>A SNP edits and PAM site sequence restrictions at the locus of interest. The initial base editing experiments relied on co-delivery of a gRNA plasmid carrying an antibiotic selection marker and a base editor plasmid (YE1-BE3 or BE3-Gam). Consequently, these experiments were limited by the selection of cells that had been successfully transfected with the gRNA plasmid through G418 selection. It is well known that plasmid size directly influences nuclear uptake efficiency during transfection (Ribeiro *et al.*, 2012; Bai *et al.*, 2017). Though we correct for the difference between plasmid size by delivering the base editors at a higher ratio than the gRNA (3:1), we were still unable to assess for cells which had undergone co-delivery of both plasmids and monitor transfection efficiency of which the base editor being so large, is a limiting factor. As we observed no single base editing during these experiments, we modified our approach to utilise the BE4-MAX GFP plasmid with a guide targeted to the RNF2 locus in order to assess the transfection efficiency and base editor expression in our iPSC line. The BE4-MAX base editor is designed with a P2A self-cleaving peptide that couples GFP production in the cell with full length BE4 production, along with a N and C terminal bipartite NLS to aid trafficking from the cytoplasm to the nucleus (Koblan *et al.*, 2018). These modifications from the third-generation editors result in enhanced gene transfer and mRNA/protein expression resulting in a greater than three-fold editing efficiency. These have been observed the biggest obstacles of editing experiments to date. Moreover, these base editors have been used to correct pathogenic SNPs disease relevant loci in diverse cell types using nucleofection with high efficiency (Koblan *et al.*, 2018). We aimed to optimised transfection efficiency using a gRNA targeted to the RNF2 locus using a gRNA sequence established in the literature as a highly editable locus by the third and fourth generation of base editors (Komor *et al.*, 2016; Grünewald *et al.*, 2019; McGrath *et al.*, 2019). Efficiencies

of 25>60% total sequence reads with a target C>T conversion (BE3 and BE4-MAX) have been reported in the literature at the RNF2 site in both human and mouse cell lines (Komor *et al.*, 2016; Koblan *et al.*, 2018). Through co-delivery of gRNF2 and BE4-MAX GFP we would improve transfection efficiency due to these modifications outlined whilst also being able to isolate cells that were expressing the gRNA- by G418 selection and cells expressing BE4-MAX through sorting GFP positive cells. Nevertheless, in our experiment low transfection efficiencies were still observed (2.15%), which is surprising as nucleofection of BE4-MAX in patient Congenital Disorders of Glycosylation (CDG) type 1 fibroblasts, murine neuroblast (Neuro-2a) cell line and HEK293T cell lines have been reported around 15, 30 and 60% GFP transfection efficiencies respectively (Koblan *et al.*, 2018). The PWP1 GFP plasmid was included as a control during this transfection due to it's large size (11,103bp) and its high transfection efficiencies reported within our lab (data not shown). However, this plasmid also showed a low transfection efficiency (2.49%) suggesting that the electroporation needs further optimisation. Delivery of DNA by electroporation is mainly dependent on three electrical parameters: electrical field, pulse width and pulse number. In these experiments we used the same parameters that had given us a high efficiency in our CRISPR/Cas9 experiments where we delivered RNP complexes, which were also reported to give high efficiency for plasmid delivery in optimised protocols from the manufacturer. However, future experiments would require full optimisation of these parameters by running a 24 well optimisation experiment where electrical field, pulse width and pulse number are investigated in our cell of interest to select for the best conditions. Following this optimisation, the next step would be to use our T354 and R361 GATA2 guides with the BE4-MAX to assess the editing efficiency of these GATA2 sequences.

Chapter 5:

Exploring the role of inflammation as a driver of disease progression during early stages of Gata2 syndromes

5.1 Introduction

The concept that there is a functional correlation between inflammation and cancer, was first established by Rudolf Virchow in 1863, where he hypothesized that infiltrated immune cells reflect the location where cancer lesions appear in inflamed tissue (Virchow, 1989). In solid cancer it is now apparent that the tumour microenvironment, in which inflammatory cells are a coordinate element, is instrumental in the neoplastic process, enabling proliferation, survival and migration of tumour cells. There is an abundance of evidence showing that tumour cells adopt many aspects of innate immune signalling molecules, to support invasion migration and metastasis (Balkwill and Mantovani, 2001; Coussens and Werb, 2002; Balkwill, Charles and Mantovani, 2005; Karin, 2006).

In haematological malignancies this theme is also being extensively studied. Chronic innate immune signalling in HSCs have been widely documented in MDS patients, along with the dysregulation of innate immune-regulated genes in HSPCs. Moreover, mutations and overexpression of pattern recognition receptors (PRR) such as TLR-4, have been documented, along with aberrant expression of down-stream effectors such as MyD88/IRAK1 and IRAK4 kinases (Barreyro, Chlon and Starczynowski, 2018a; Matos, Magalhães and Rauh, 2020). Furthermore, activation of the innate immune pathway by PRR are correlated with an increasing risk of MDS development (Barreyro, Chlon and Starczynowski, 2018b).

There has also been growing evidence that inflammation plays a key role in the evolution of Clonal haematopoiesis of indeterminant potential (CHIP). In 2015, Steensma and colleagues first characterised CHIP to describe the precursor state of haematological neoplasia. In this state, individuals have an expansion of a blood cell clone, derived from a single HSC, containing a somatic mutation in a driver gene associated with leukaemia (Steensma *et al.*, 2015a). However, these individuals do not meet the WHO diagnostic criteria for haematological malignancy and mostly have normal complete blood counts. These mutated clones are expressed at a variant allele frequency (VAF) of 0.02 (2%) and are documented in 10-15% of the population ~70 years of age. This increases to 30% in individuals 85 years of age and older and presents a 0.5-1% annual risk of haematological malignancy (Steensma *et al.*, 2015b; Steensma, 2018b). Furthermore, CHIP has also shown to increase risk of myocardial infarction, stroke and cardiovascular morbidity nevertheless, there are individuals where CHIP results in no adverse consequence to overall health (Jaiswal *et al.*, 2014b, 2017; Bazeley, Morales and Tang, 2020; Fawaz *et al.*, 2021). It has been suggested that inflammation resulting from clonal haematopoiesis (CH), may lead to an increased proliferation pressure on HSCs or progenitor cells, which in-turn contributes to genomic

instability and consequently clonal evolution (Steensma, 2018a). Mutations in TET2, the second most frequently observed mutation in CHIP, has been extensively studied in the last few years. *TET2* has a well-established role in the control of inflammation via suppression of IL-6 and has been documented in AML, MDS and various inflammatory related diseases (Montagner *et al.*, 2016; Q. Shen *et al.*, 2018; Lio, Yuita and Rao, 2019; Kunimoto and Nakajima, 2021). Work conducted by Zhang *et al* (2015) outlined that during inflammation resolution, Tet2 is responsible for selectively mediating repression of IL-6 transcription in innate myeloid cells (DC and macrophages). Using an endotoxin shock model in myeloid cell specific *Tet2*-deficient mice stimulated with LPS challenge, the authors revealed an increased production in IL-6 with reports of severe tissue damage and diffuse inflammation in the lungs of mice. The repression mechanism of *Tet2* was highlighted to be achieved through *Tet2* recruitment of histone deacetylase *Hdac2* at the *Il6* promoter, in order to prevent persistent transcriptional activation at the chromatin level for inflammation resolution. Loss of *Tet2* during LPS stimulation has also been shown to increase Il-1b, Il-6 and Arginase 1 (Arg1) mRNA levels in BM derived macrophages *in vitro* (Cull *et al.*, 2017a). In accordance with these findings Cai *et., al* (2018) showed that TET2-deficent HSPCs exhibit strong proliferation advantages and a myeloid-bias in response to LPS induced inflammation. These innate myeloid cells were shown to produce high levels of pro-inflammatory cytokines such as IL-6, to maintain survival whilst also supressing apoptosis (Cai *et al.*, 2018). This evidence suggest a positive feedback loop between *Tet2*-deficent HSPCs and their progeny during the response to inflammatory-induced cytokine production (Q. Zhang *et al.*, 2015). This adaptation of an abnormal innate immune response in TET2-mutatant HSPCs has been observed not only in the haematopoietic system but also in the peripheral tissues and cardiovascular system (Q. Zhang *et al.*, 2015; Fuster *et al.*, 2017; Sano *et al.*, 2018). Taken together, these studies suggest that the loss of TET-2 function in myeloid cells may alter the immune environment in patients with *TET2*-mutant CHIP and *TET-2* mutant MDS and AML patients (Cull *et al.*, 2017a).

Clinical evidence also suggests that inflammation may contribute to disease severity in germline *RUNX1* mutations, responsible for the FPD syndrome and predisposition to AML (***described in section 1.3.1***). All three Runx proteins play vital roles in immune system development and function and RUNX1 mutations are correlated with an increase in immune signalling pathways (Djuretic, Cruz-Guilloty and Rao, 2009; Bellissimo and Speck, 2017). *Runx1* specific deletion in lung epithelium has been shown to increase susceptibility to LPS-induced acute lung injury *in vivo* via NFκB signalling (Tang *et al.*, 2017) and T cell specific deletion has been reported to cause hyperactivation of $CD4^+$ T cells, resulting in severe lung inflammation which evolved into systemic inflammatory disease (Djuretic, Cruz-Guilloty and Rao, 2009). RUNX1-deficient BM neutrophils have been shown to upregulate NFκB activity

and overactive inflammatory pathways gaining competitive advantage under chronic inflammation (Bellissimo *et al.*, 2020). In line with these findings, unpublished research from the Majeti group have demonstrated that whilst loss of *RUNX1* in $CD34^+$ HSPCs initially presents a competitive disadvantage, as shown by the decreased proliferation, cell cycle arrest and a decrease in engraftment potential, gene set enrichment analysis (GSEA) of transcriptional data revealed an upregulation of NF-κB-mediated inflammatory pathways. Furthermore, they detected an upregulation of the IL-3 receptor (CD123) in *RUNX1* KO cells, which drove expansion *in vivo* and *in vitro*. Together, these finding elude to how *RUNX1* mutations can initially behave in a deleterious manner however, under inflammatory environmental conditions confer a selective advantage to HSPCs and may potentially lead to disease progression (Fan *et al.*, 2020).

Similarly, patients with *GATA2* hereditary mutations suffer with cytopenias in many of the haematopoietic lineages associated with vital components of innate and adaptive immunity (Bigley *et al.*, 2011; Dickinson *et al.*, 2014; Spinner *et al.*, 2014). Clinical evidence has shown that some patients have an impairment of whole blood cytokine responses, in particular presenting with defects in the production or signalling of type I cytokines, resulting in predisposition to otherwise weakly pathogenic mycobacteria (Bigley *et al.*, 2011). Defects in the production of IL-12 from phagocytic effectors have been thought to lead to the inability of adaptive immune response, initiation through ineffective activation and differentiation of T cells into T-helper cells, ultimately resulting in reduced IFN-γ production and ineffective control of intracellular pathogens (Bigley *et al.*, 2011; Ramirez-Alejo and Santos-Argumedo, 2014). These defects likely describe how these patients become highly susceptible to recurrent bacterial and viral infections. T-cells are generally well preserved in peripheral pool due to their independence from the BM compartment, however, inversion of the CD4:CD8 ratio in these patients have also been documented (Bigley *et al.*, 2011). This alteration is an indication of an immune risk phenotype resulting from declining CD4 helper cells during chronic antigen stimulation, causing expansion of CD8 memory cells (McBride and Striker, 2017). This also implies an overall reduction in the T cell repertoire in these patients causing further susceptibility to future infections (Vinh *et al.*, 2010; Ostergaard *et al.*, 2011; West *et al.*, 2014). Taken together, the literature proposes a premature aging phenotype in these GATA2 immunodeficient patients where the immune system adapts in order to mount a sufficient immune response due to the presence of chronic antigen stimulation. This phenomenon also referred to as ‘inflamm-aging’ describes the functional decline of the haematopoietic system that is correlated to the aging population (Verovskaya, Dellorusso and Passegué, 2019). During the process of aging HSCs lose their regenerative capacity leading; to immune

senescence, anaemia and a bias toward myeloid output, along with systemic accumulation of proinflammatory cytokines from these myeloid cells (Pietras, 2017; Broxmeyer *et al.*, 2020; Zhang *et al.*, 2020). These events lead to an increased risk of age-related clonal haematopoiesis (ARCH), autoimmunity and haematological malignancy. Furthermore, single cell RNA-sequencing performed from BM of eight GATA2 patients with evidence of MDS, revealed genes of the immune system where the most abundantly dysregulated (Wu *et al.*, 2020). In support of this, RNA sequencing data generated from our own lab in young adult *Gata2* heterozygous mice during steady state, has revealed upregulation of IFN-γ and TLR signalling pathways (Abdellfattah 2021-*unpublished*)(Bothmer *et al.*, 2010).

Taking into consideration the distinct roles that inflammation plays in solid cancers, precursor stages and haematological malignancies throughout the literature, we hypothesise that systemic inflammation induced by chronic antigen stimulation from autoimmunity and recurrent infections in GATA2 patients, leads to an inflammatory BM microenvironment, providing a positive selection pressure for further mutation acquisition and disease progression.

5.2 Aims & Objectives

The specific aims and objectives for this chapter are:

1. To study the impact of reduced *GATA2* expression in CB CD34$^+$ HSPCs via shRNA-mediated knockdown during myeloid differentiation in the presence of LPS stimulation.

2. To assess the replating potential of *Gata2* KD CB CD34$^+$ HSPCs during CFC serial replating in the presence of LPS stimulation.

3. To investigate the impact of *Gata2* heterozygosity during chronic inflammation using the Vav-iCre conditional mouse model by pIpC stimulation for a period of seven days *in vivo*.

4. To assess the replating potential of *Gata2* haploinsufficent cells by CFC serial replating *in vitro.*

5.3.1 Validation of shRNA vectors against human GATA2 in CB HPSCs.

To decipher the impact of dysregulation of GATA2 in HSPCs during an inflammatory stimulus, $CD34^+$ CB cells were isolated and transduced using a bicistronic lentiviral vector system encoding shRNA against human GATA2 with a GFP reporter. Knock down efficiency of *GATA2* was validated from three independent shRNA constructs targeting *GATA2*; sh31, sh32 and sh34 in comparison to a shRNA construct containing a scramble sequence used as a control (shSC). After three days post transduction a 30-35% transduction efficiency was achieved in $CD34^+$ CB. These cells were FACS sorted for GFP^+CD34^+ cells and were validated for knock down efficiency **(Figure 5.1 A and B).** At the RNA level the knock-down efficiency was ~60% for construct 32 and ~90% for sh31 and sh34 **(Figure 5.1 C)**. As it is well known that the relationship between RNA levels and expression of the resulting protein do not always correlate due to posttranslational modification and other factors (Vogel and Marcotte, 2012), we also validated at the protein level by western blot. After normalisation of *GAPDH* control loading we found a knock-down efficiency of ~80% for sh31 and ~90% for sh32 and sh34 **(Figure 5.1 D)**. In further experiments, sh32 was carried forward and is hereafter referred to as GATA2 KD along with shSC as a control.

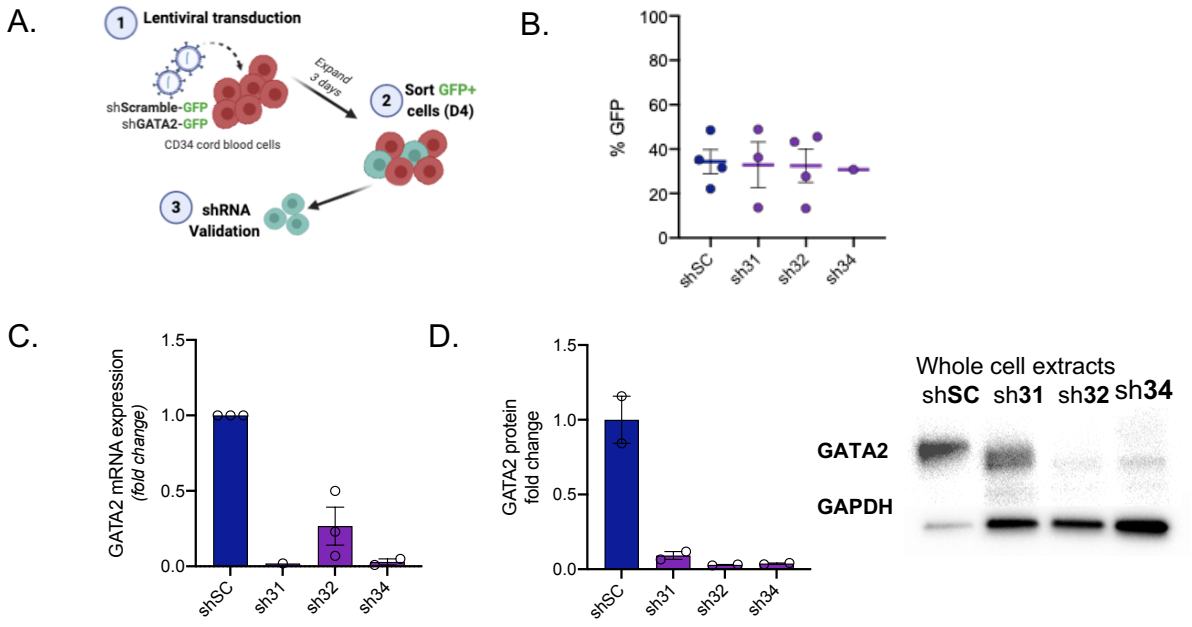


Figure 5.1: Knockdown validation of human short hairpin constructs against GATA2. (A) Human $CD34^+$ cord blood cells were transduced with lentiviruses encoding a short hairpin against human *GATA2* or Scramble sequence control with a GFP reporter. Three days post transduction cells were FACS sorted and knockdown efficiency validated. **(B)**Transduction efficiency of shRNA in $CD34^+$ CB

cells, followed by validation by **(C)** qPCR (n=3) and **(D)** by western blot (n=2) *GAPDH* was used as a housekeeping gene. Data shown are ± SEM. Image created with BioRender.com.

5.3.2 LPS stimulation in GATA2 KD cells results in an increased progenitor output.

To investigate whether GATA2 deficient HSPC gain a competitive advantage for disease initiation during a continuous inflammatory pressure, we transduced $CD34^+$ cells with either shSC or sh32 knockdown constructs and plated $CD34^+GFP^+$ cells into CFC assays in the presence of 10µg/ml LPS **(Figure 5.2).**

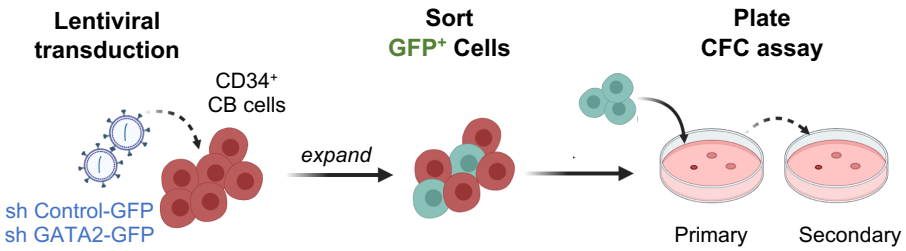


Figure 5.2: Colony forming cell assay of GATA2-KD or control $CD34^+$ CB cells stimulated with LPS. $CD34^+$ cord blood cells were transduced with lentiviruses encoding a short hairpin against human *GATA2* with a GFP reporter or a scramble sequence control with a GFP reporter. Cells were expanded in culture for 3 days, FACS sorted for GFP^+ cells and 500 cells plated into a CFC assay for 10 days treated with 10µg/ml LPS or PBS. Image created with BioRender.com.

Treated and untreated control cells formed a comparable number of CFUs overall, with no changes to CFU type. In-line with our previous data, we observed a significant reduction in CFU numbers between control groups and GATA2 KD control HSPCs due to a significant decrease in BFU-E formation. This reduction of CFU output was partially rescued in the presence of LPS treatment due to an expansion of BFU-E formation and a non-significant expansion of combined CFU-M and GM colonies **(Figure 5.3 A)**. Moreover, we observed a non-significant increase in replating capacity of CFU-E and BFU-E colonies derived from GATA2 KD HSPCs, which were slightly reduced with treatment of LPS **(Figure 5.3 B)**. Nevertheless, this replating capacity was not sustained as tertiary CFC assays produced no colonies within any experimental groups (Data not shown).

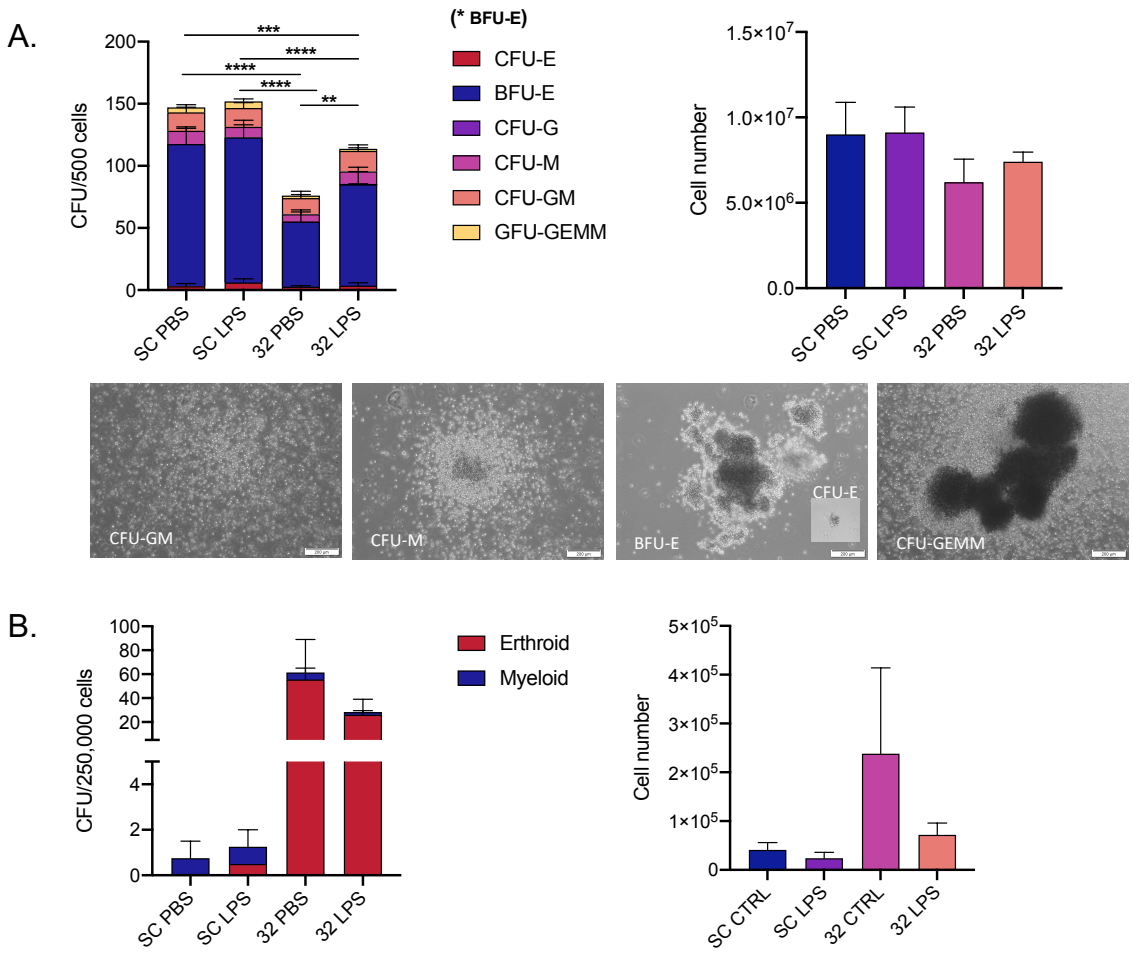


Figure 5.3: LPS stimulation in GATA2 KD cells results in an increased progenitor output. (A) Number and classification of colony type at day 10, along with representative CFC images. Statistical analysis: two-way ANOVA (n=4). **(B)** Number of colonies from secondary plating and total cell number (n=2). Data shown are ± SEM. P values: ** p<0.01, ***p<0.001, **** p<0.0001.

5.3.3 GATA2 KD cells have reduced late phase apoptosis

Myeloid cells are granulocytic and phagocytic innate leukocytes which are critical for mounting a successful immune response. Upon recognition of pathogen-associated molecular patterns (PAMPs) they initiate a signalling cascade which ultimately result in the production of pro inflammatory cytokines in order to control host infections (Iqbal, Fisher and Greaves, 2016). To assess whether a continuous inflammatory stimulus combined with *GATA2* dysregulation impacts the differentiation of HSPCs towards the myeloid lineage, we sorted $CD34^{+}GDP^{+}$ shSC (control) or sh32 (*GATA2* KD) HSPCs and seeded them into liquid culture in the presence of myeloid expansion cytokines for two weeks, with or without LPS stimulation **(Figure 5.4)**. During myeloid expansion the total numbers of *GATA2* KD cells, irrespective of LPS treatment, were significantly decreased at day 7 of culture and at day 14, control and *GATA2* KD cells remained decreased **(Figure 5.5 A)**. To understand whether these cells were

being lost through an apoptotic or necrotic process we conducted an Annexin V assay on cells collected from day 7 cultures. We found a significant reduction of *GATA2* KD cells undergoing apoptosis, specifically in the late apoptotic phase, regardless of treatment condition **(Figure 5.5 B)**. To assess whether treated *GATA2* KD cells had altered proinflammatory signalling we assessed TLR-4 expression in myeloid cells at day 7. TLR-4 is a PRR that becomes activated during infection through recognition of LPS located in tissues or the blood stream, subsequently triggering a pro-inflammatory response to the invading bacteria. Activation occurs with the help of LPS-binding protein (LBP) and CD14 along with MD-2 protein which is stably associated with the extracellular fragment of the receptor (Ciesielska, Matyjek and Kwiatkowska, 2020). Prolonged activation of TLR-4 has been linked with several human hereditary, autoimmune diseases and cancer (Korneev *et al.*, 2017; Barreyro, Chlon and Starczynowski, 2018a). We found no significant differences in the abundance of cells expressing TLR-4 ligand, or to the overall MFI between experimental groups **(Figure 5.5 C)**.

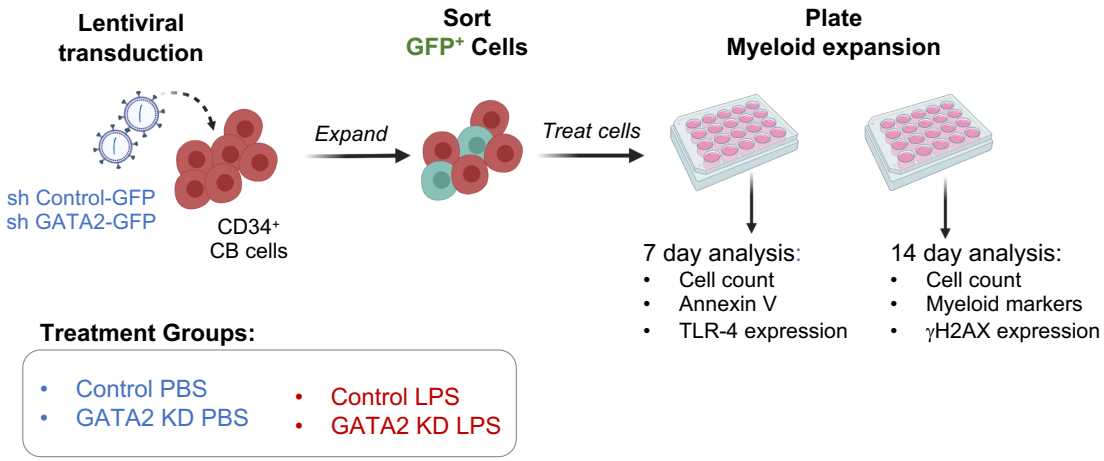


Figure 5.4: Myeloid differentiation of GATA2 KD or control CD34⁺ CB cells stimulated with LPS. CD34⁺ cord blood cells were transduced with lentiviruses encoding a short hairpin against human *GATA2* with a GFP reporter or a Scramble sequence control with a GFP reporter. Cells were expanded in culture for 3 days, FACS sorted for GFP⁺ cells and seeded into myeloid expansion liquid culture medium containing 1µg/ml LPS or equal volume of PBS. Cells were harvested and analyzed on days 7 and 14. Image created with BioRender.com.

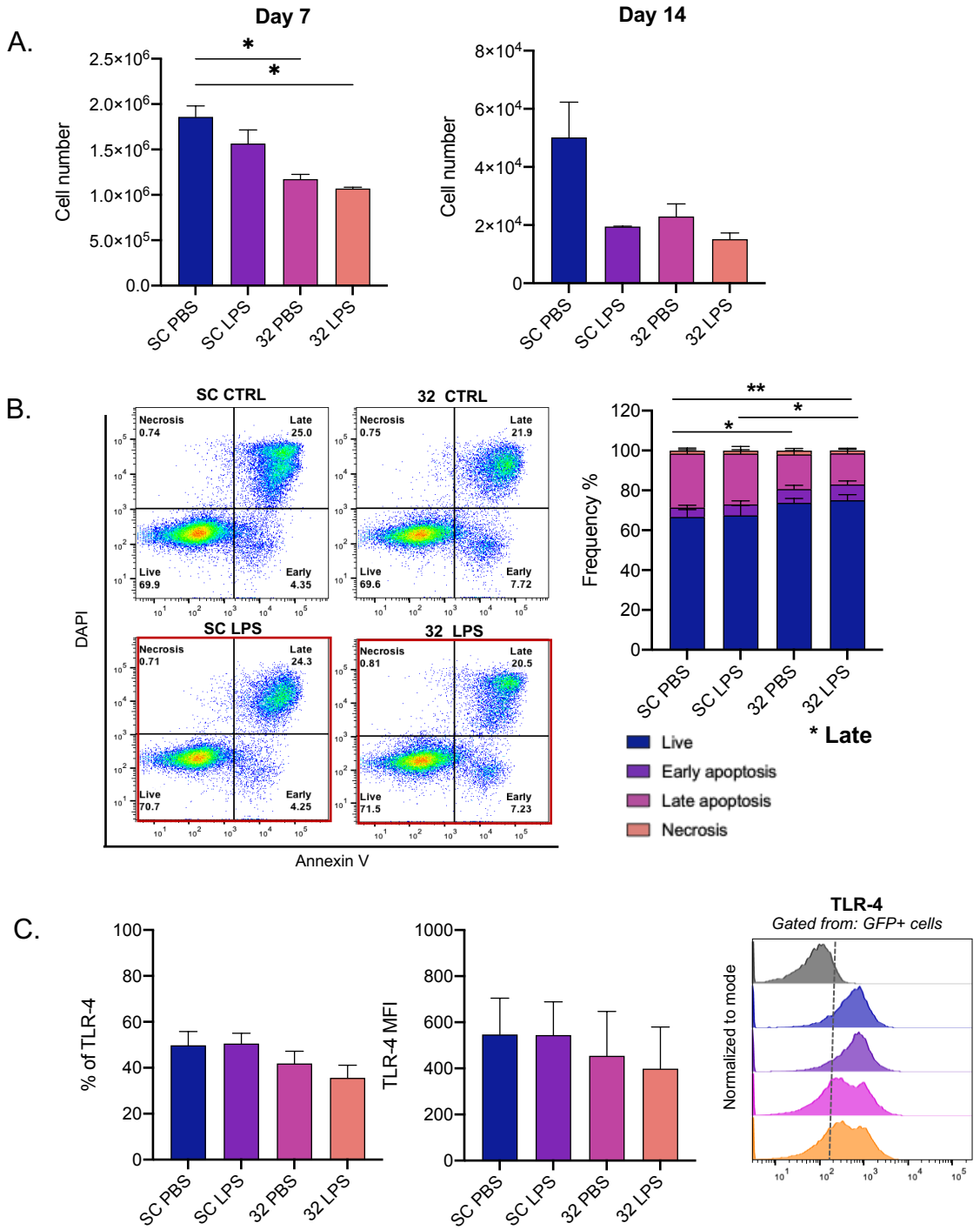


Figure 5.5: LPS stimulation in GATA2 KD cells results in reduced cell number during myeloid differentiation and a GATA2 mediated effect on late apoptosis. (A) Cell number at day 7 and 14 time points (n=4). Statistical analysis: one-way ANOVA. **(B)** Frequency of AnnexinV$^+$ cells at day 7 (n=4). Statistical analysis: two-way ANOVA. Representative FACS plots showing apoptotic phase. **(C)** Frequency of TLR-4$^+$ cells and representative FACS histogram showing TLR-4 protein expression compared to unstained control (Grey). MFI of TLR-4 (n=4). Data shown are ± SEM. P-values: *p<0.05, **p<0.01.

5.3.4 CD14 is highly expressed in *GATA2* KD cells stimulated with LPS.

After two weeks in culture cells were harvested and assessed for markers of myeloid differentiation. We observed a significant decrease in abundance and MFI of $CD14^+$ from GATA2 KD cells compared to control, independent of LPS stimulation. Interestingly, upon careful examination of the CD14 histogram plots, we noticed that within the $CD14^+$ fraction there was a clear shift in CD14 expression to the right during LPS stimulation, which was significant in *GATA2* KD cells **(Figure 5.6 A)**. Finally, as there is a correlation between DNA damage and inflammation due to the increase in reactive oxygen and nitrogen species being produced, we assessed our cultures for signs of DNA damage (Chatzinikolaou, Karakasilioti and Garinis, 2014; Pérez-Figueroa *et al.*, 2021). We identified no differences in both the abundance and MFI of cells expressing γ-H2AX **(Figure 5.6 B)**.

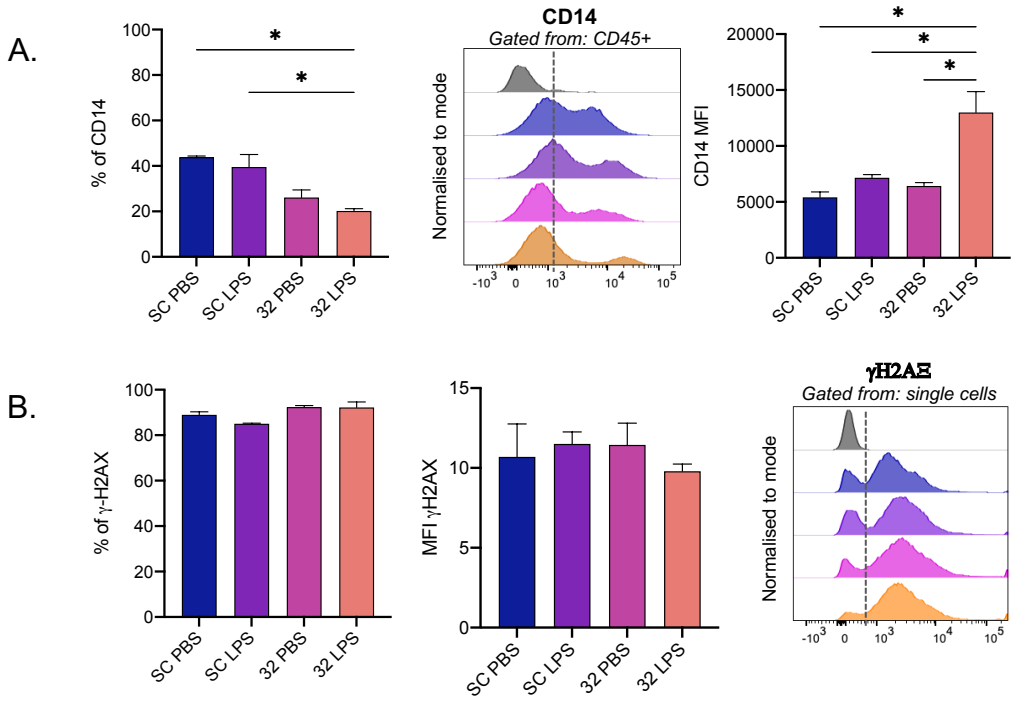


Figure 5.6: CD14 is highly expressed in *GATA2* KD cells stimulated with LPS. GATA2 KD or control cells were seeded into myeloid expansion liquid culture medium containing 1µg/ml LPS or equal volume of PBS. Cells were harvested and analyzed at day 14. **(A)** Frequency of $CD14^+$, representative FACS histogram showing CD14 expression and MFI of positive CD14 cells (left to right) (n=2). Statistical analysis: one way ANOVA. **(C)** Frequency and MFI of gH2AX representative FACS histogram showing γH2AX expression at day 14 (n=2). Data shown are ± SEM. *p<0.05

5.3.5 pIpC stimulation in *Gata2* heterozygous mice results in decreased myeloid output in the BM.

In order to explore the impact of *Gata2* mediated inflammation *in vivo*, we employed the VaviCre mouse model to specifically delete one *Gata2* allele in the haematopoietic

compartment using the Vav promoter. These *Gata2*$^{+/fl}$ mice were injected with double-stranded RNA analog polyinosinic:polycytidylic acid (pIpC) every other day for seven days to induce a high IFN type 1 response (IFNα/β), reminiscent of a chronic viral insult as described by the Passague group (Pietras *et al.*, 2014; Jalbert and Pietras, 2018) **(Figure 5.7 A)**. The day after the last injection mice were sacrificed and detailed immunophenotyping was carried out of hemopoietic compartments in the BM and SP from HSC to mature progeny. pIpC is recognised by endosomal Toll-like receptor 3 (TLR-3) which subsequently activates TF interferon regulatory factor3 (IRF3) via the adapter protein Toll-IL-1 receptor domain-containing adapter (TRIF/TICAM-1), ultimately resulting to the production of type-1 IFNs. Moreover, pIpC can also lead to recruitment of TNF receptor-associated factor-6 (TRAF6) and subsequent activation of TF NFκB and AP-1 ultimately triggering inflammatory cytokines and chemokines. Lastly, pIpC can also activate cytosolic RNA helicases retinoic acid-inducible protein (RIG-1 and melanoma differentiation-associate gene 5 (MDA-5) (Alexopoulou *et al.*, 2001; Lee *et al.*, 2006; Kawai and Akira, 2008).

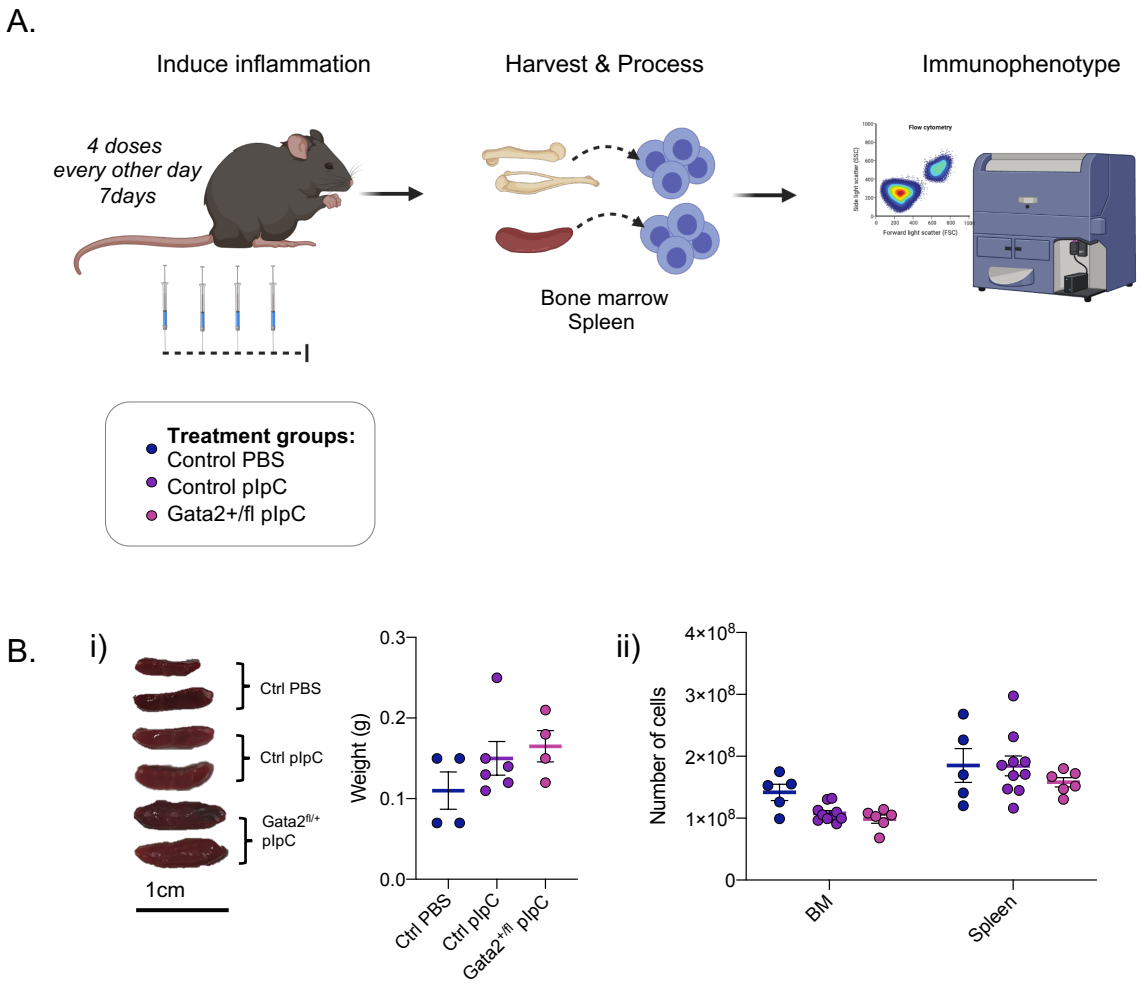


Figure 5.7: pIpC stimulation in Gata2 heterozygous mice results in splenomegaly. (A) *Gata2*$^{+/fl}$;VaviCre or control mice were administered 10mg/kg pIpC every other day for 7 days. At day 7 the mice were culled the bone marrow and spleen harvested and immunophenotyped. **(B)(i)** Representative images of spleen size and weights (n=4/6/4). **(ii)**Bone marrow and spleen cellularity. Data shown are ± SEM. Each dot represents one individual mouse (n=5/10/6). Image created with BioRender.com.

5.3.5 pIpC stimulation in Gata2 heterozygous mice results in splenomegaly.

Splenomegaly was evident in *Gata2*-het mice although the weight and cellularity of the spleen were unchanged, along with BM cellularity **(Figure 5.7 B**). Histological analysis of spleen is ongoing. However, preliminary analysis suggests the formation of granulomas in *Gata2* heterozygous animals which may explain the Gata2 mediated splenomegaly following pIpC administration (data not shown). Granulomas are clusters of immune cells in affected tissues and have been documented in several autoinflammatory disease along with GATA2 immunodeficient patients as a result of chronic inflammation due to recurrent infections (Collin, Dickinson and Bigley, 2015a; Donadieu *et al.*, 2018). Looking in the different mature cell lineages within the BM, we observed a significant reduction in both the frequency and

abundance of B-cells between PBS and pIpC control mice **(Figure 5.8 B-D)** and no effects to the T cell subsets $CD4^+$ and $CD8^+$. However, a trend towards a reduction of the CD4:CD8 ratio between *Gata2*-het and pIpC control mice was also observed **(Figure 5.8 E)**.

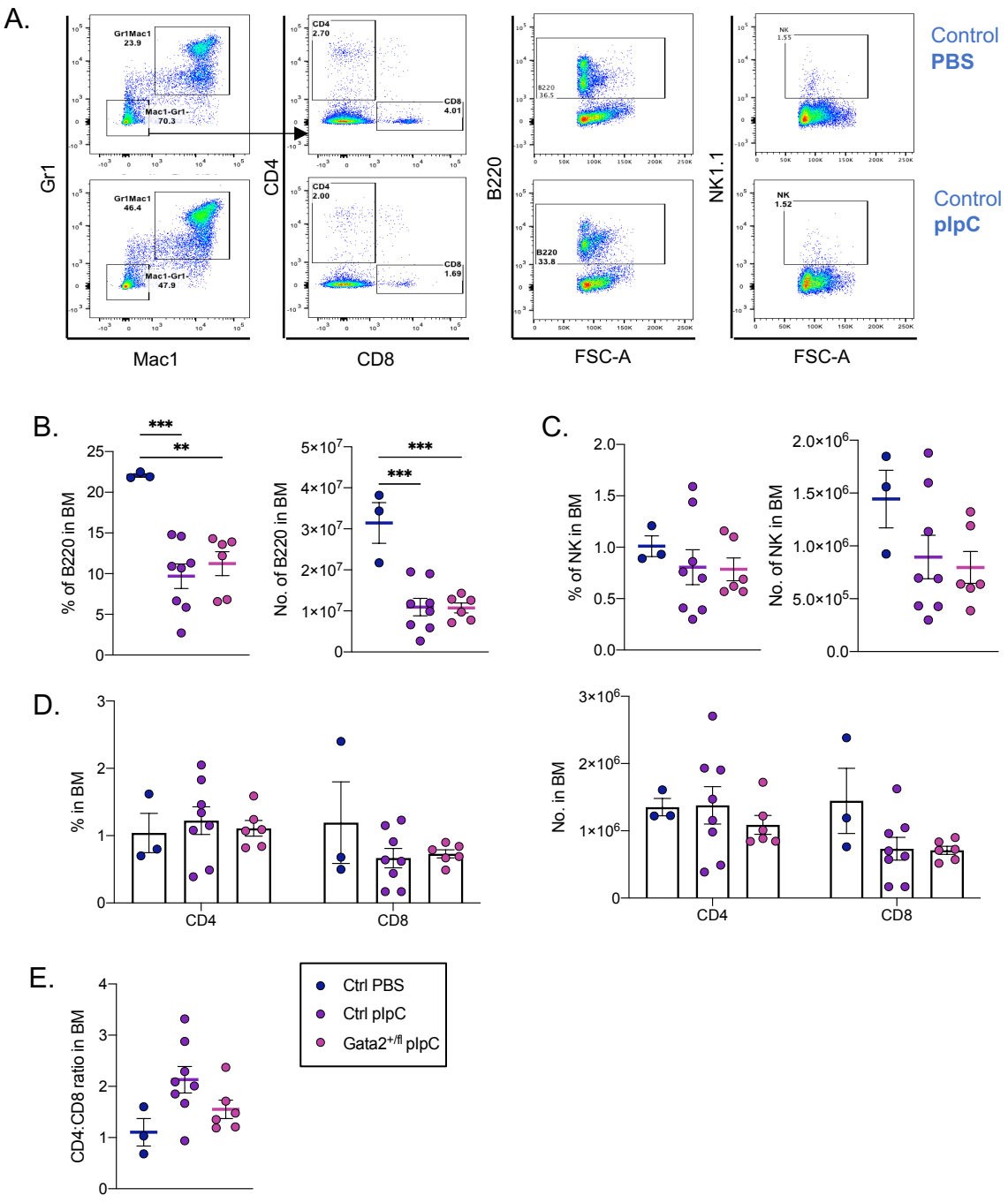


Figure 5.8: *Gata2* heterozygosity during chronic pIpC stimulation in mice has no effect on mature lymphoid cells in the BM. (A) Gating strategy used: Cells were initially gated from single cells, followed by live cells as previously described. Mature lymphoid cells were gated from the $Gr1^{-}Mac^{-}$ population. Representative FACS plots from control mouse treated with PBS and Gata2 control moused treated with pIpC. **(B-D)** Frequency and cellularity of indicated populations within the BM (n=3/6). **(E)** CD4:CD8 ratio in the BM. Data shown are ± SEM, each dot represents one individual mouse (n=3/8/6). Statistical analysis: one-way ANOVA. **p<0.01, ***p<0.001.

We observed a significant increase in the frequency of immature myeloid precursors marked by $Gr1^{+}Mac1^{+}$ in control mice whilst interestingly the abundance of these cells were significantly decreased *Gata2*-het mice compared to pIpC treated control mice **(Figure 5.9 B)**. As it is well known that myeloid cells play an important role during inflammation, we further investigated this reduction in myeloid cells by looking in more detail at the monocyte subsets in the bone marrow with inflammatory functions (Shi and Pamer, 2011; Yang *et al.*, 2014). Whilst there was a significant decrease in the frequency and abundance of patrolling monocytes marked by $CD115^{+}Ly6C^{high}$ between PBS and pIpC treated mice there were no overall *Gata2* mediated changes during inflammatory challenge *in vivo* **(Figure 5.9 C)**.

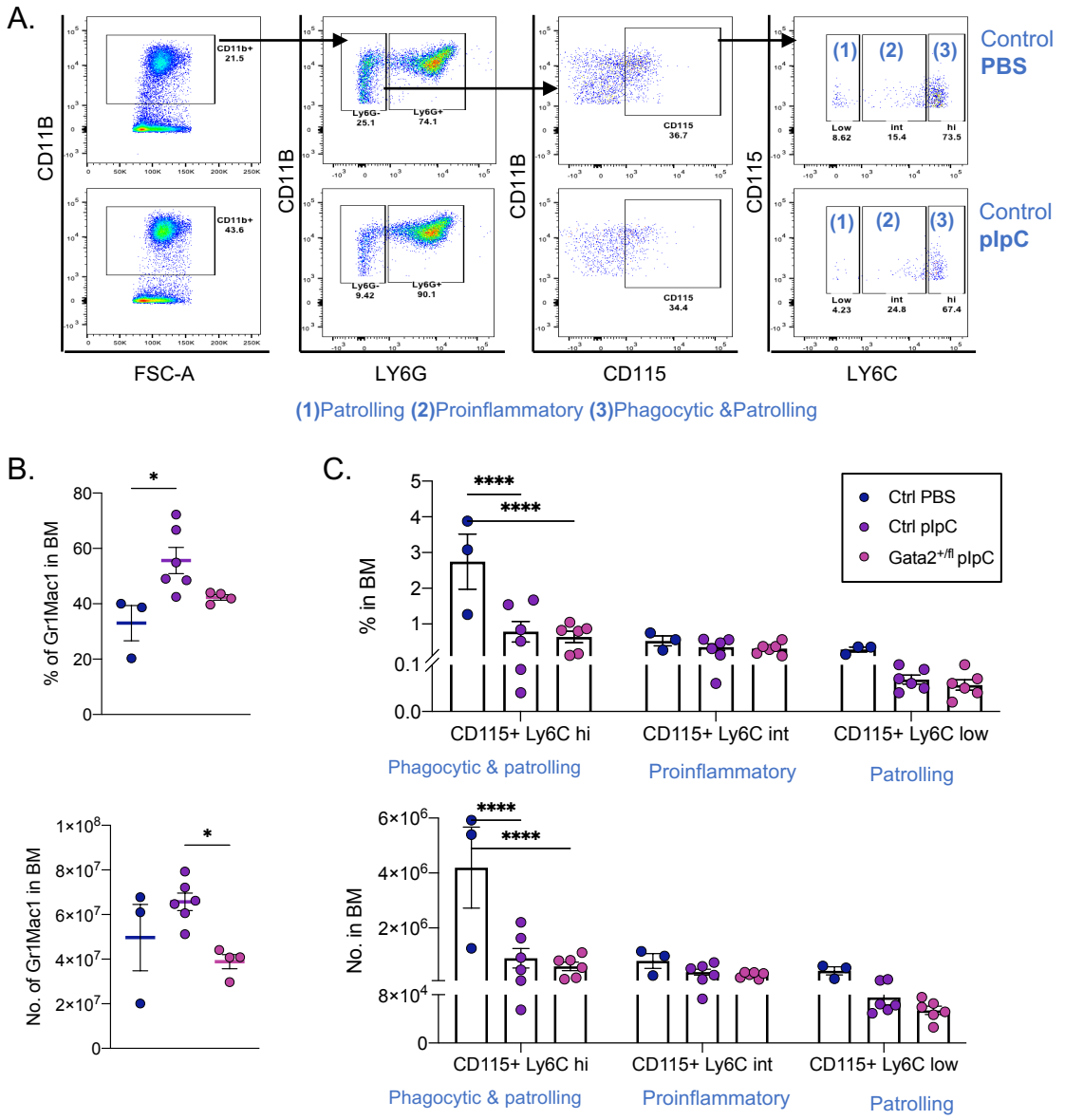


Figure 5.9: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on multipotent monocyte subsets in the BM. (A) Gating strategy used: Cells were initially gated from single cells, followed by live cells as previously described. Monocyte subsets were gated from $CD11b^{+}Ly6G^{-}CD115^{+}$, this population can be subdivided into three populations: $Ly6C^{high}$ (monocytes with phagocytosis and pro-inflammatory functions), $Ly6C^{intermediate}$ (monocytes with proinflammatory function) and $Ly6C^{low}$ (patrolling and tissue repair). Representative FACS plots from control mouse treated with PBS and Gata2 control moused treated with pIpC. **(B)** Frequency and cellularity of $Gr1^{+}Mac1^{+}$ cells in the BM (n= 3/6/4). **(C)** Frequency and cellularity of indicated monocyte populations within the BM (n=3/6). Data shown are ± SEM, each dot represents one individual mouse. Statistical analysis: one-way ANOVA. *p<0.05**** and p<0.0001.

Next, we investigated any changes in myeloerythroid restricted precursor cells which are derived from multipotent haematopoietic cells. These progenitor subsets include early bipotent progenitors for erythroid/megakaryocyte lineages 'PreMegE', early monopotent erythroid; 'PreCFU-E' and 'proEry+CFU-E' and megakaryocyte progenitors 'MkP', along with primitive granulocyte/macrophage progenitors Pre-GM and GMP **(Figure 5.10)** (Pronk and Bryder, 2018). We observed a significant decrease in the frequency and abundance of PreGM cells between control animals, however no differences were seen in the GMP compartment **(Figures 5.11 A and B)**. All immature erythroid compartments were significantly reduced between control groups though no significant differences were observed between Gata2-het and control pIpC treated mice **(Figures 5.11 C-F)**.

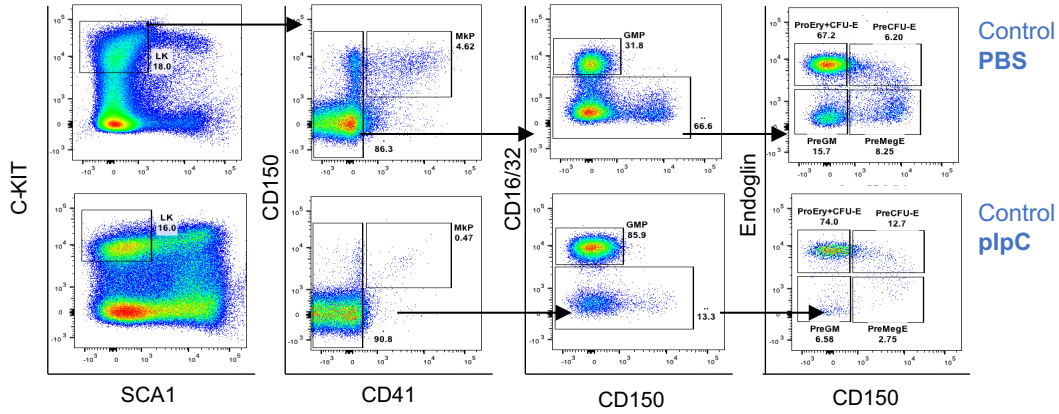


Figure 5.10: Gating strategy for immature myeloerythroid cell lineages in the bone marrow during inflammation. Lin$^-$Sca1$^+$ were gated and subsequently c-Kit$^-$Sca1$^+$ cells to define the LK compartment. CD150$^+$CD41$^+$ defines a population of megakaryocyte progenitors (MkP), CD41$^-$ CD16/32$^+$CD150$^-$ defines granulocyte/macrophage progenitor population (GMP), the CD16/32$^{-/low}$ fraction can be subdivided into four populations: pre granulocyte/macrophage *pre-GM* (Endoglin$^-$ CD150$^-$), pre-megakaryocyte/erythroid PreMegE (Endoglin$^-$CD150$^+$), pre-colony forming unit-erythrocytes Pre CFU-E (Endoglin$^+$CD150$^+$) and pro-erythroid colony forming unit-erythrocytes pro-EryCFU-E (Endoglin$^+$CD150$^-$). Representative FACS plots from a control mouse treated with PBS and *Gata2* control moused treated with pIpC.

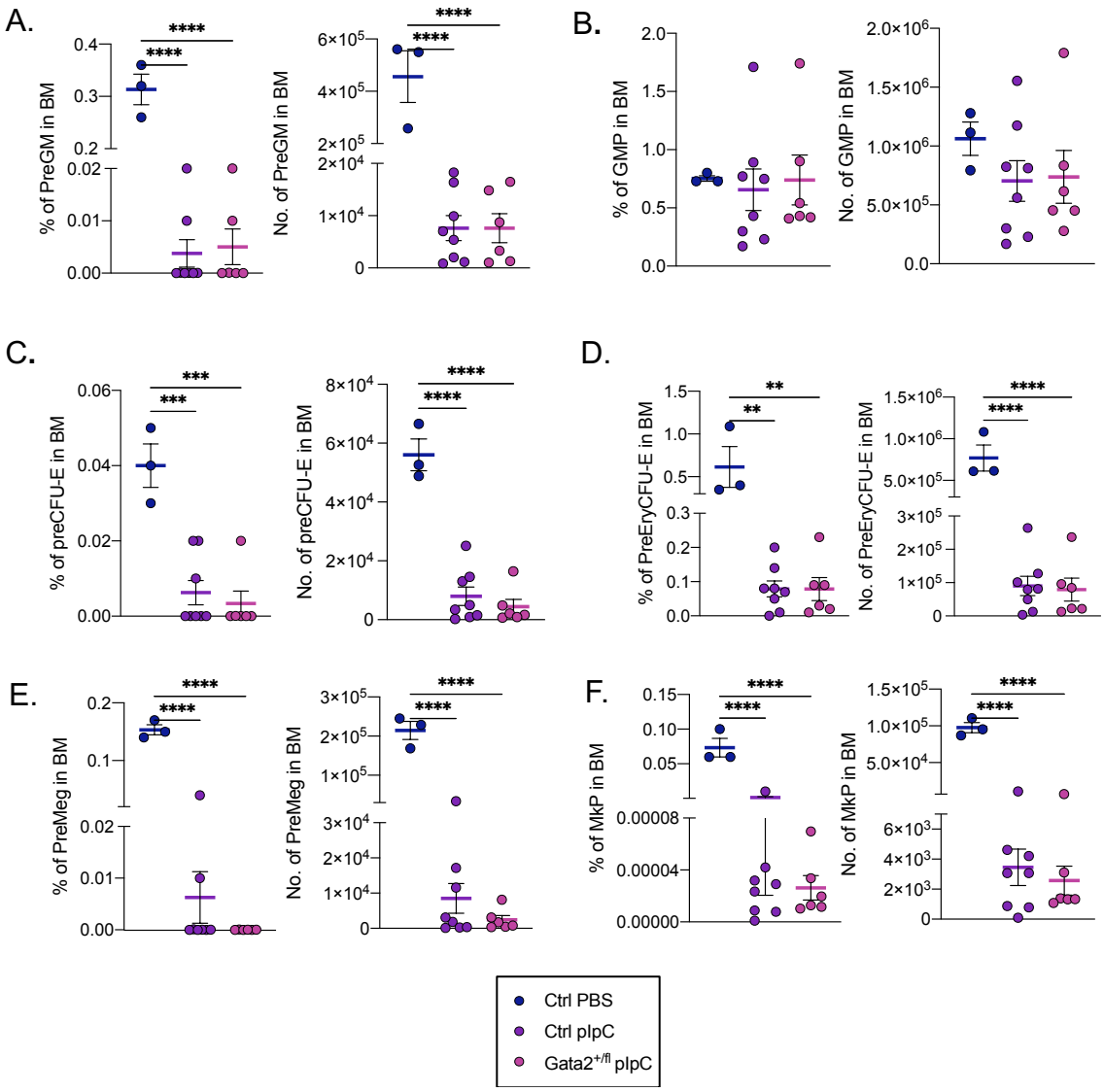


Figure 5.11: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on immature myeloerythroid cell lineages in the bone marrow. (A-F) Frequency and cellularity of indicated immature myeloerythroid cell lineages within the bone marrow. Data shown are ± SEM, each dot represents one individual mouse (n=3/8/6). Statistical analysis: one-way ANOVA. **p<0.01, ***p<0.001 and ****p<0.0001.

Finally, total BM cells were harvested and seeded into a colony forming assays *in vitro* to functionally assess whether pIpC treatment had altered their capacity to form colonies and any potential for clonal expansion **(Figure 5.12 A)**. We found no differences in progenitor output between all groups in primary CFC assays. In the first replating of colonies, we observe a non-significant expansion of myeloid colonies between pIpC treated control and *Gata2*-het cells however, this was not sustained in the secondary re-plating **(Figure 5.12 D)**.

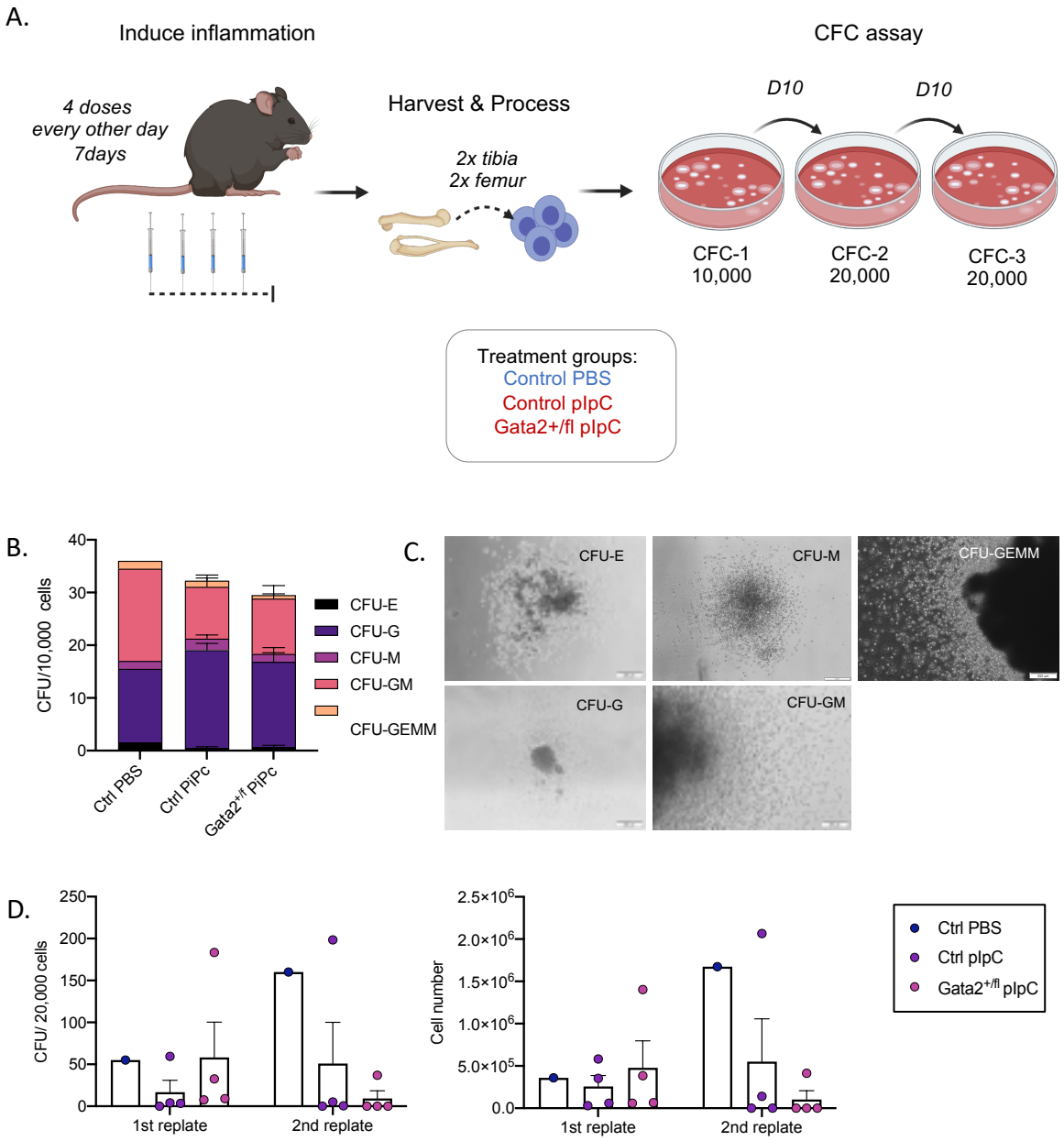


Figure 5.12: CFC and replating assays from total BM cells from pIpC stimulated *Gata2* heterozygous mice. (A) Experimental design; total BM cells were cultured in CFC assays *in vitro* and subsequently plated into secondary and tertiary CFC assays. **(B)** Number and classification of colony type at day 10 (n=1/4) **(C)** Representative images of individual colony types. **(D)** Colony and cell number from secondary and tertiary CFC plating. Each dot represents one individual mouse (n=1/4). Image created with BioRender.com.

5.3.6 pIpC exposure in *Gata2* heterozygous mice has no effect on stem and progenitor output in the BM.

Taking into consideration previous work conducted in our lab by Dr Ali Abdelfattah also using the VaviCre model to investigate the effects of *Gata2* haploinsufficiency which showed a

significant decrease in stem and progenitor cells **(Figure 5.22 and 5.23)**, in line with Rodrigues et al., (2008) and with the well documented notion that inflammatory signals regulate HSC fate (Pietras, 2017), we investigated the impact of chronic inflammation on HSC and early MPP populations in the BM. As expected, we observed an expansion in the frequency and number of LSK cells in mice stimulated with pIpC typically observed as a result of the reactivation of Sca1 expressing myeloid progenitors due to IFN-1 exposure, complemented by a significant decrease in the CLP frequency and abundance between PBS and pIpC induced mice **(Figure 5.13 and Figure 5.14 A-D)** (Pietras *et al.*, 2014). Moreover, we also observed a significant increase in the frequency of myeloid biased MPP progenitors (MPP2 and MPP3) and a significant reduction in the abundance of lymphoid biased MPP4 progenitors between control mice **(Figure 5.14 C)**. Surprisingly, there was no differences observed in the HSC compartment, with the exception of long-term HSC abundance being significantly reduced between PBS and pIpC treated mice **(Figure 5.14 E)**. Taken together, no overall Gata2 mediated changes were observed in the stem and progenitor populations during an inflammatory stimulus.

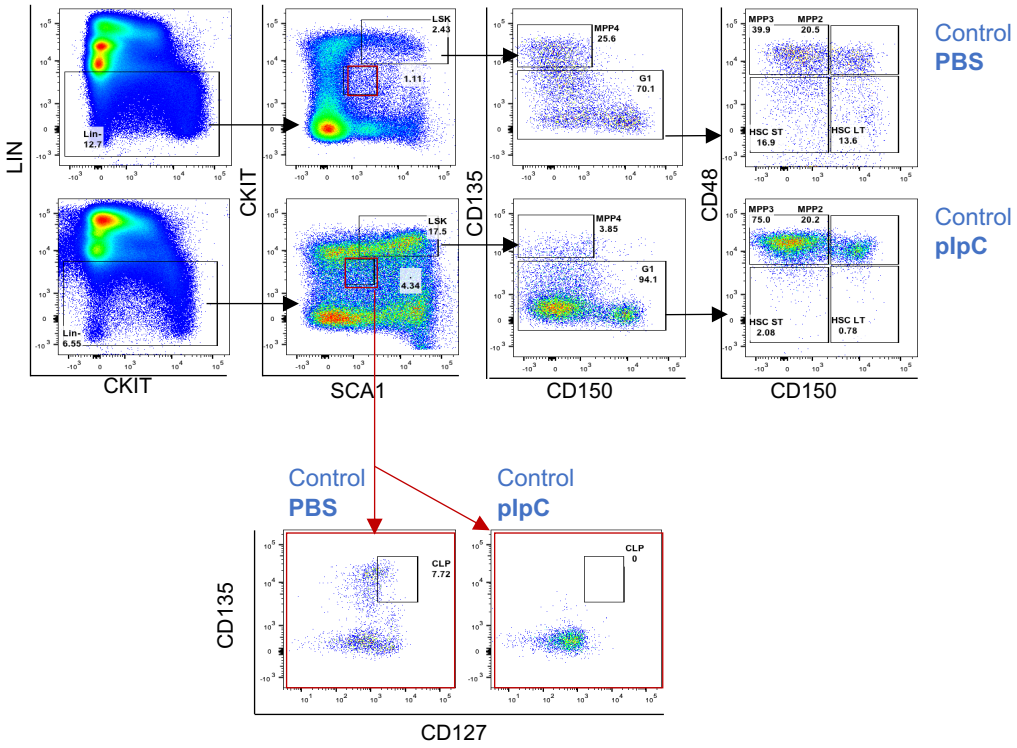


Figure 5.13: Gating strategy for multipotent progenitor and common lymphoid progenitor compartments in the bone marrow during inflammation. c-Kit^{+}Sca1^{+} cells were gated from Lin^{-} Sca1^{+} to define the LSK compartment. From the LSK, we can identify MPP4 (CD135^{+}CD150^{-}) and from the CD135^{-} compartment we can isolate MPP3 (CD48^{+}CD150^{+}), MPP2 (CD48^{+}CD150^{-}), ST-HSC (CD48^{-}CD150^{-}) and LT-HSC (CD48^{-}CD150^{+}) populations. A c-Kit^{mid}Sca1^{mid} population of cells is used to define the CLP compartment. Representative FACS plots from a control mouse treated with PBS and *Gata2* control moused treated with pIpC evident by the clear shift in the in Sca1 positive cells, resulting in increased frequency of the LSK population.

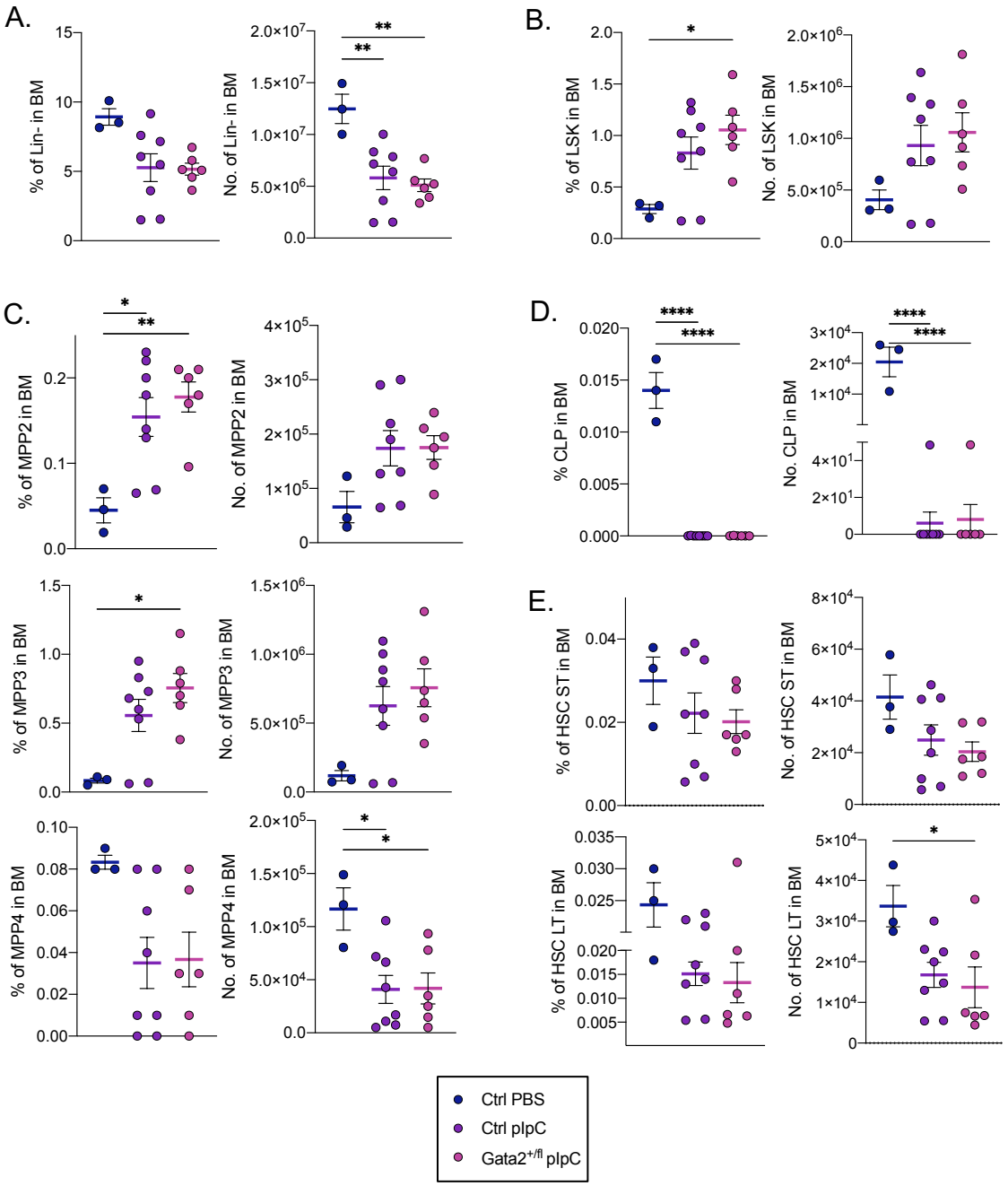


Figure 5.14: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on multipotent progenitor and common lymphoid progenitor populations in the BM. (A-E) Frequency and cellularity of indicated multipotent progenitor populations within the bone marrow. Data shown are ± SEM, each dot represents one individual mouse (n=3/8/6). Statistical analysis: one-way ANOVA. *p<0.05, **p<0.01 and ****p<0.0001.

5.3.7 pIpC exposure in *Gata2* heterozygous mice has no effect mature cells in the spleen.

Though the BM is typically the major source of circulating monocytes, under stress conditions extramedullary haematopoiesis can also provide a substantial contribution of these cells (Iqbal, Fisher and Greaves, 2016). Studies have highlighted that HSPCs migrate from the BM to specialised niches in the spleen where they proliferate and give rise to monocytes that enter the circulation to the site of inflammation (Leuschner *et al.*, 2012). In light of the marked increase in spleen size observed in our *Gata2* heterozygous mice, we carried out detailed immunophenotyping in this organ. Surprisingly, we found no overall *Gata2* mediated changes during inflammation in the frequency or abundance of mature lymphoid **(Figure 5.15)** or myeloid cells **(Figure 5:16 A)**, along with no changes to the more specific monocyte subsets **(Figure 5.16 C)**. We further investigated into stem and progenitor populations within the spleen however, again there were no *Gata2* mediated changes during inflammation *in* vivo **(Figure 5.17 A-D, Figure 5.18 A and 5:19 A-G)**. Finally, as monocytes are mobilised from the BM and are recruited to sites of inflammation, we also investigated mature populations in the PB. In line with these kinetics, we observed a non-significant increase in the frequency of $Gr1^{+}Mac1^{+}$ cells circulating in the blood between control animals treated with PBS and pIpC. Interesting, this was further increased in *Gata2* mice treated with pIpC **(Figure 5.20 A-B)**.

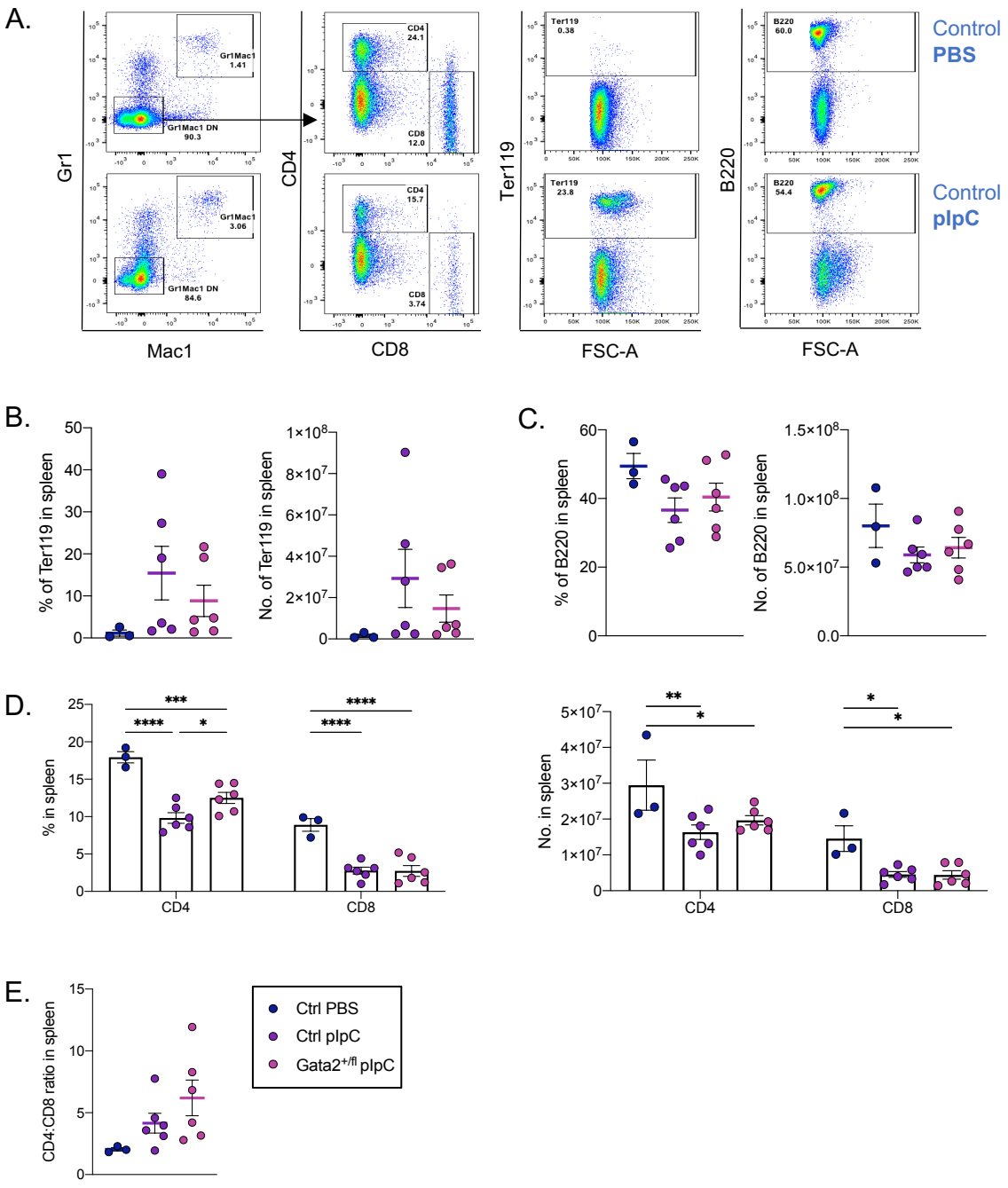


Figure 5.15: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on mature populations in the spleen (A) Representative FACS plots defining the gating strategy for mature cell populations in the spleen control mouse treated with PBS and *Gata2* control moused treated with pIpC. **(B-E)** Frequency and cellularity of indicated populations within the spleen (n=3/6). **(E)** CD4:CD8 ratio in the BM. Data shown are ± SEM, each dot represents one individual mouse (n=3/6). *p<0.001, ***p<0.01 and ****p<0.0001

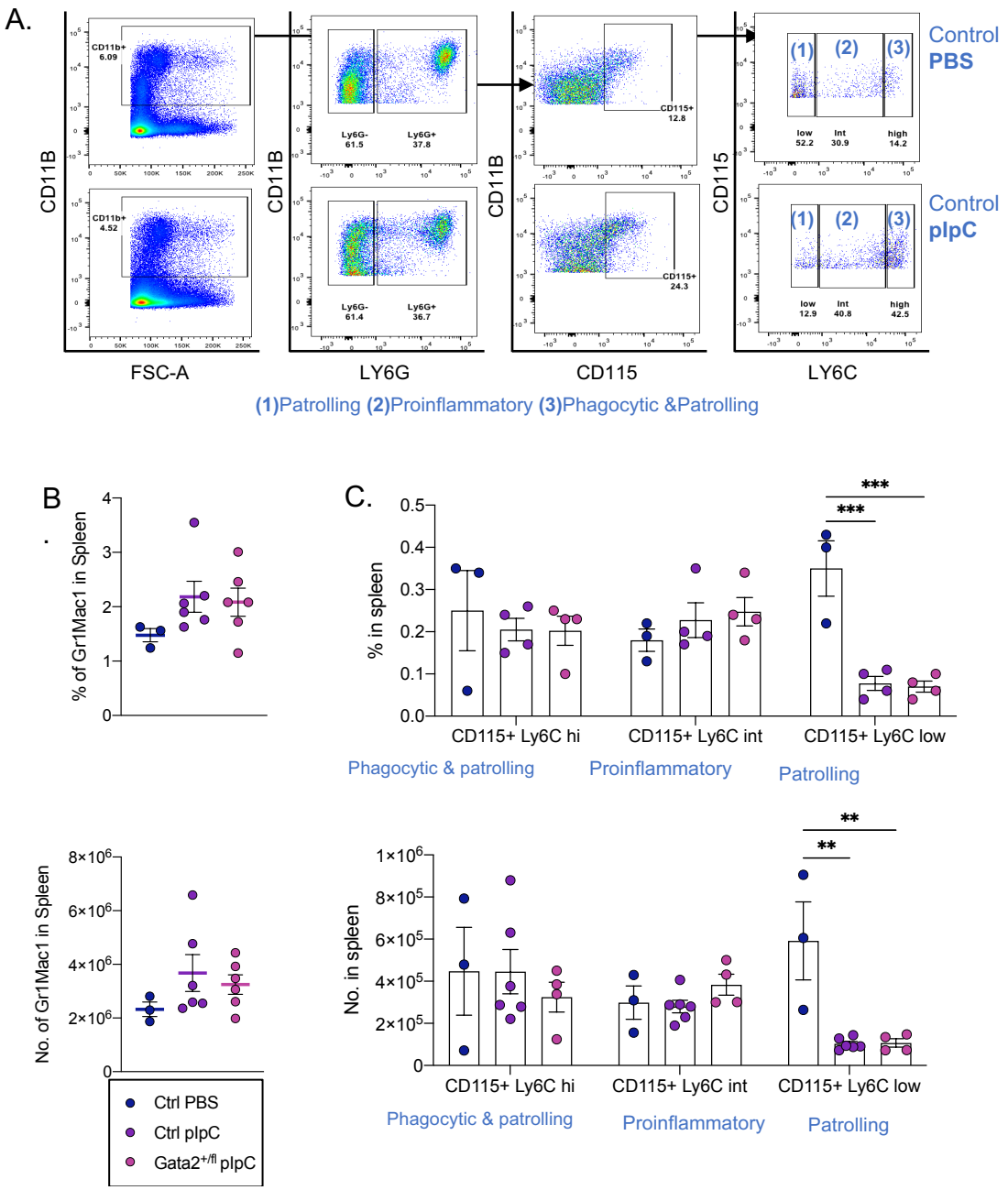


Figure 5.16: *Gata2* heterozygosity during chronic pIpC exposure in mice has no effect on multipotent monocyte subsets in the Spleen. (A) Representative FACS plots from control mouse treated with PBS and Gata2 control moused treated with pIpC. **(B)** Frequency and cellularity of indicated populations within the spleen (n=3/6). **(C)**Frequency and cellularity of Gr1+Mac1+ cells in the spleen (n= 3/6/5). Data shown are ± SEM, each dot represents one individual mouse. Statistical analysis: Two-way ANOVA. **p<0.01 ***p<0.001.

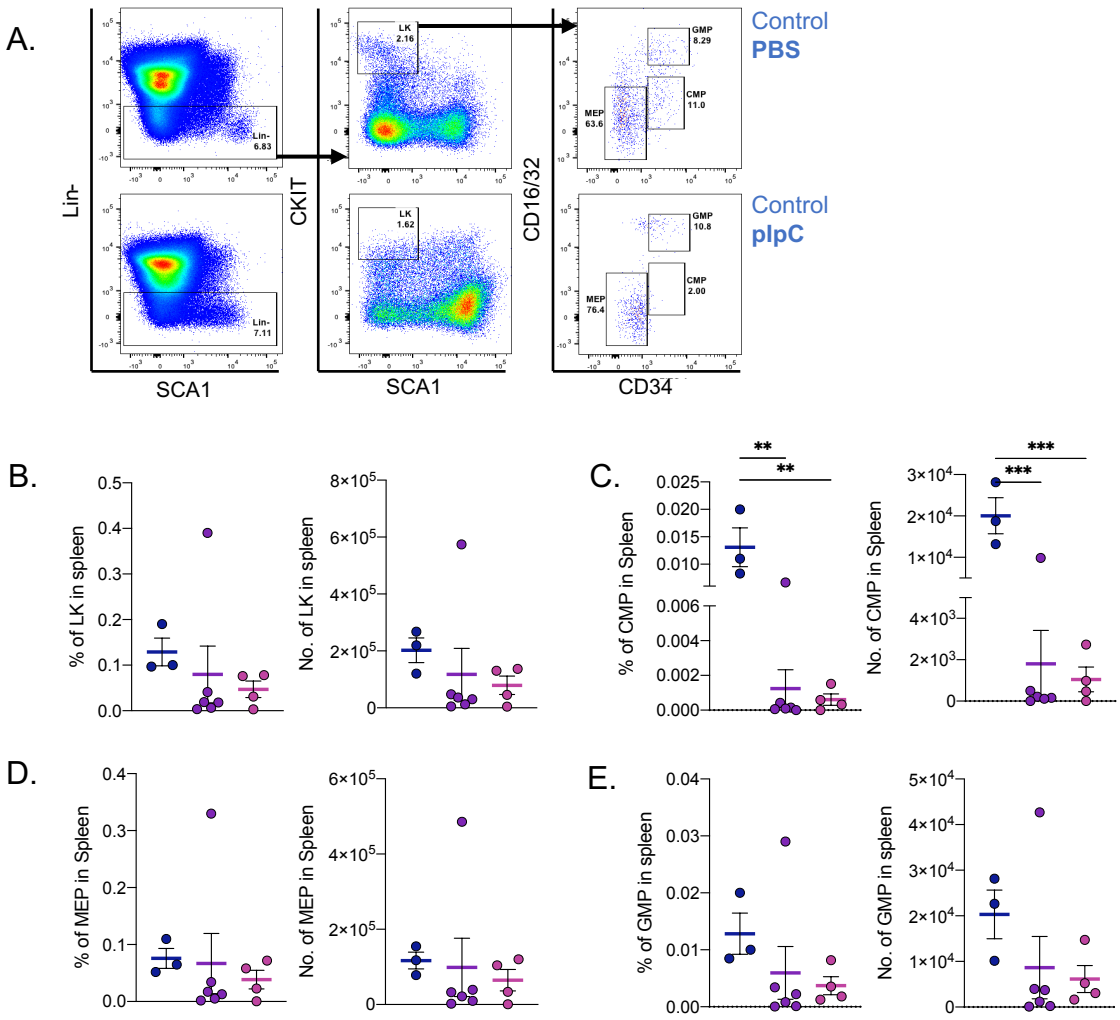


Figure 5.17 Chronic pIpC exposure in *Gata2* heterozygous mice has no impact on progenitor populations in the spleen. (A) Representative FACS plots defining the gating strategy for progenitor populations in the spleen control mouse treated with PBS and *Gata2* control moused treated with pIpC. **(B-E)** Frequency and cellularity of indicated progenitor populations within the spleen. Data shown are ± SEM, each dot represents one individual mouse (n=3/6/4). Statistical analysis: Two-way ANOVA. **p<0.01 ***p<0.001.

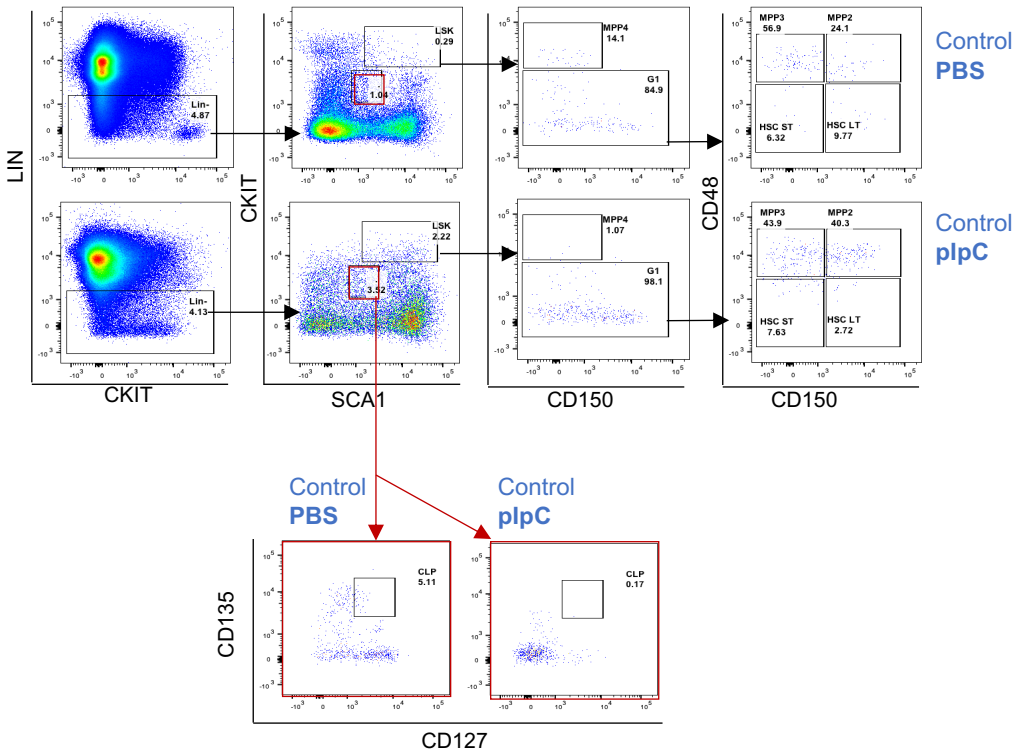


Figure 5:18: Gating strategy for stem and progenitor cells in the spleen. Representative FACS plots from control mouse treated with PBS and Gata2 control moused treated with pIpC.

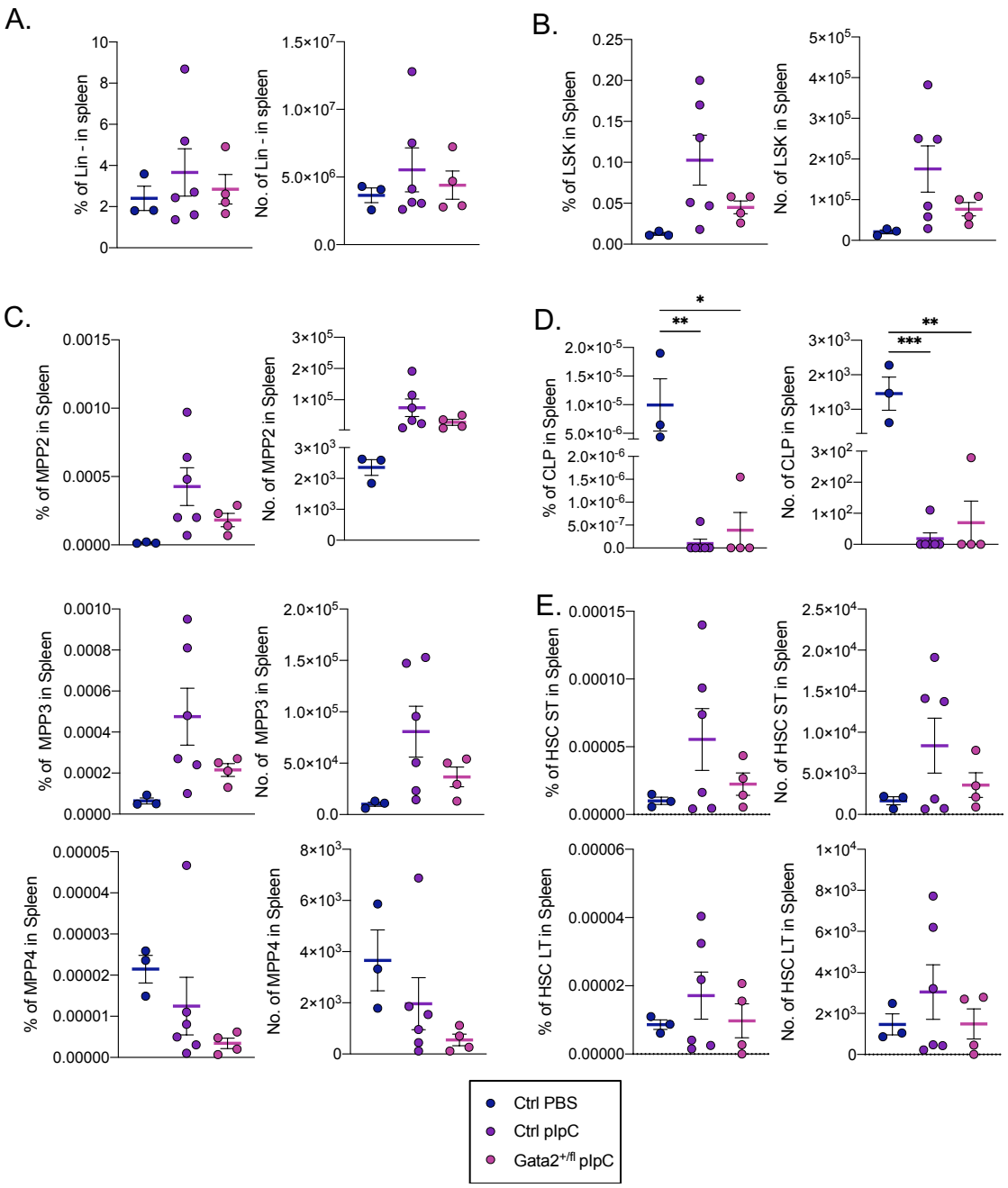


Figure 5.19: Chronic pIpC exposure in *Gata2* heterozygous mice has no effect on multipotent progenitor and common lymphoid progenitor populations in the spleen. (A) Representative FACS plots from control mouse treated with PBS and Gata2 control moused treated with pIpC. **(A-E)** Frequency and cellularity of indicated multipotent progenitor populations within the bone marrow. Data shown are ± SEM, each dot represents one individual mouse (n=3/6/4). Statistical analysis: One-way ANOVA. *p<0.05, **p<0.01, ***p<0.001.

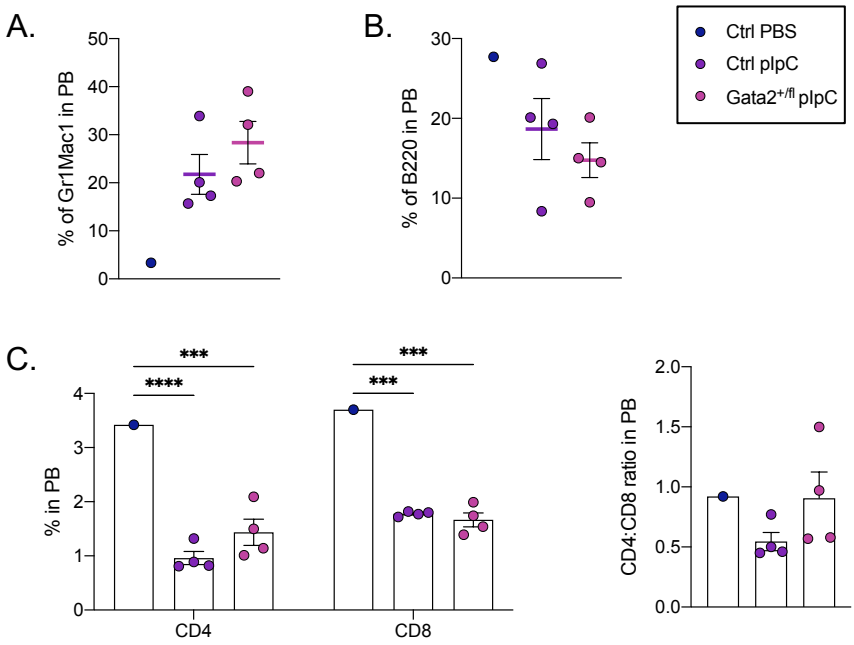


Figure 5.20: *Gata2* heterozygosity during chronic pIpC exposure in mice leads to an increase in myeloid precursors in the PB. (A-C) Frequency and cellularity of indicated populations within the peripheral blood. (C) CD4:CD8 ratios of the PB. Data shown are ± SEM, each dot represents one individual mouse (n=1/4). Statistical analysis: Two-way ANOVA. ***p<0.001.

5.3.8 Chronic pIpC stimulation in *Gata2* haploinsufficient mice does not impact the DNA damage response.

There is a strong relationship between chronic inflammation and increased DNA damage as a result of increased levels of intracellular reactive oxygen species (ROS) and naturally produced genotoxic metabolites from effector cells (Chatzinikolaou, Karakasilioti and Garinis, 2014). Furthermore, RNA sequencing data conducted by Ali Abdelfattah in age matched *Gata2* heterozygous mice in our laboratory, revealed an upregulation of DNA repair pathways, further supported by a significant increase in the MFI of γ-H2AX in the BM of HSPCs (Abdelfattah, 2021-*unpublished*). We therefore finally assessed whether these mice under IFN-1 stimulation have increased DNA damage response. We observed no difference in the expression of γ-H2AX phosphorylation in HSPCs of pIpC treated controls and *Gata2* treated mice **(Figure 5.21)**.

Taken together, our *in vivo* data where we investigated the effects of a seven-day chronic inflammatory stimulus in *Gata2* haploinsufficient mice showed a reduction in myeloid cells within the BM with no other *Gata2* mediated changes in the BM or spleen.

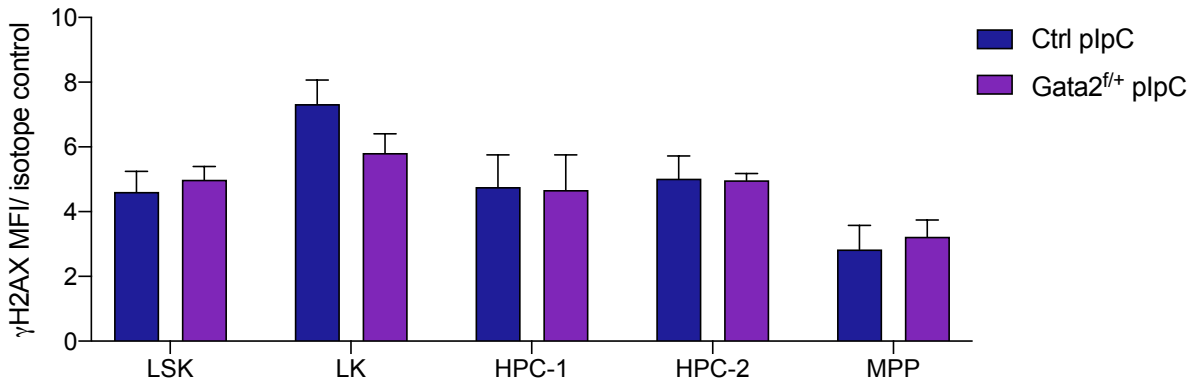


Figure 5.21: *Gata2* heterozygosity during chronic pIpC exposure in mice shows no effect of phosphorylation of γ-H2AX in the BM. MFI of γ-H2AX expression in indicated populations within the bone marrow. Data shown are ± SEM, (n=4).

5.4 Discussion

In this chapter we investigated whether inflammation drives disease progression during early stages of GATA2 syndromes. To test this hypothesis, we applied two different approaches under an inflammatory stimulus; *in vitro* knockdown of *GATA2* in CB HSPCs along with *in vivo* specific deletion of a single *Gata2* allele in the haematopoietic compartment using the Vav promoter.

A key part of the innate immune system, monocytes play both a direct and indirect role in antimicrobial immunity via phagocytosis and cytokine production, along with the regulation of innate and adaptive immune cells (Cormican and Griffin, 2020). In our *GATA2* KD experiments during myeloid differentiation, we observed that *GATA2* KD cells produce less $CD14^{+}$ myeloid cells irrespective of LPS treatment which was further supported by a significant reduction of $Gr1^{+}Mac1^{+}$ cells in the BM of *Gata2*-het mice treated with pIpC. Intriguingly, upon further investigation of our *in vitro* data we found a specific increase in CD14 expression within the $CD14^{+}$ positive fraction in *GATA2* KD cells stimulated with LPS. This is noteworthy finding as although CD14 is a myeloid marker antigen, it is also a component of the LPS receptor complex that induces signal transduction through TLR-4, triggering the synthesis and release of proinflammatory chemokine IP10 and cytokines (TNF-I, IL-6 and IL-1) (Guha and Mackman, 2001; Zamani *et al.*, 2013). The monocyte pool in both human and mice are composed of three subsets, in human this is based on their expression of CD14 and the CD16 receptor (FcγRIII); classical ($CD14^{hi}CD16^{-}$), intermediate ($CD14^{hi}CD16^{+}$) and nonclassical ($CD14^{low}CD16^{hi}$) (Wong *et al.*, 2011a; Sugimoto *et al.*, 2015). This could suggest that *GATA2* KD cells during a short period of inflammation, differentiate into classical or intermediate subpopulations of monocytes which highly express CD14, whilst control cells irrespective of LPS and control GATA2 KD cells express CD14 at lower levels, reminiscent of the more terminally differentiated non-classical subpopulation. These populations have a variety of different functions regulating innate and adaptive immune responses to pathogens, outlined in key studies through a range of gene expression analysis and *in vitro* functional assays (Cros *et al.*, 2010; Wong *et al.*, 2011b; Zawada *et al.*, 2011). Classical and non-classical monocyte subpopulations have been shown to be primed for phagocytosis, innate sensing or immune responses and migration, intermediate for antigen presentation, cytokine secretion, apoptosis regulation and differentiation. This could suggest a differential biological response within an inflammatory environment that favours immune sensing and activity in cells with perturbed *GATA2* expression. Nevertheless, our data is preliminary and further analysis would be required to assess these points with more detail. Immunophenotyping of these monocyte subpopulations based on the expression of CD16 together with CD14 and additional assays

to assess the functionality of these cells in terms of their cytokine production, phagocytic activity and antigen presentation would be required to draw further conclusions (Skrzeczyńska-Moncznik *et al.*, 2008). In contrast to the increase in CD14 positivity found during the myeloid differentiation culture, TLR-4 expression was not specifically increased in *GATA2* KD cells and remained similar between control and *GATA2* KD cells irrespective of LPS treatment. Binding of TLR-4 through LPS-BP and CD14 decreases TLR-4 expression with duration of LPS stimulation due its internalisation along with the CD14 molecule which are degraded and recycled back to the surface of the cell membrane (Ciesielska, Matyjek and Kwiatkowska, 2020). This dynamic process and is cell type specific with TLR-4 expression being at its highest at the cell membrane in some cells at 10 minutes compared to 2 hours in others (Akbar *et al.*, 2016; Ali *et al.*, 2019). Therefore, the time at which TLR-4 expression is analysed is vital to generate a true representation of effects to its expression in the context of *GATA2* KD in these cells and requires further investigation.

The kinetic events observed in our *in vivo* modelling of inflammation between PBS and pIpC mice has been well documented throughout the literature. It is well established that IFN signalling promotes the cycling of quiescent HSC's to induce an expansion of LSK, MPP2 and MPP3 at the expense of CLP populations in order to regenerate the demand on the haematopoietic system during inflammation (Baldridge *et al.*, 2010; De Bruin, Voermans and Nolte, 2014; Pietras *et al.*, 2014; Jalbert and Pietras, 2018). Therefore, we can be confident that pIpC stimulation in our hands recapitulate a chronic inflammatory stimulus as documented in these experiments. During our *in vivo* modelling we found no prominent changes to the BM compartment of *Gata2* het mice during inflammation. Nevertheless, after just seven days of pIpC treatment Gata2 heterozygous mice displayed clear signs of splenomegaly, though no changes in WBC count or abundance of stem or progenitor cells were found in the spleen.

Despite the lack of overall impact of *Gata2* during an inflammatory setting, it is important to recognise although pIpC stimulation over a period of seven days constitutes a chronic inflammatory response, as outlined by the Passague lab (Pietras *et al.*, 2014), this is still a relatively short window of chronic inflammation. GATA2 immunodeficient patients suffer with recurrent infections and therefore go through repetitive infection and recovery cycles (Bigley *et al.*, 2011; Spinner *et al.*, 2014; Sologuren *et al.*, 2018). To more closely model this kinetics in our mouse model, ideally, repetitive pulsing and recovery of pIpC treatment over a longer experimental period would more closely resemble these recurring infections, which may be essential in producing a more accurate phenotype which mirrors these immunodeficient patients. Alternatively, in order to model *Gata2* heterozygosity in a more physiological relevant inflammatory setting *Gata2* het mice could be exposed with mycobacteria strains as this is

one of the most prevalent infections observed in GATA2 immunodeficient patients (Hsu *et al.*, 2011, 2013a; Koegel *et al.*, 2016; Mendes-De-Almeida *et al.*, 2019).

Another important factor to take into consideration is that this model specifically deletes *Gata2* from the haematopoietic compartment. Therefore, we are unable to model the interactions of *Gata2* het cells in the haematopoietic system with their niche components. Both *Gata2* and *Gata3* have been found to be highly expressed in preadipocytes, controlling their differentiation to adipocytes through binding of regulatory elements of the PPARγ promoter (Tong *et al.*, 2000), along with their physical interaction with CEBPα and CEBPβ which are positive regulators of adipogenesis (Tong *et al.*, 2005; Xu *et al.*, 2009). Moreover, *Gata2* has also been highlighted as a key regulator of MSC function as observed through *in vitro* and *in vivo* deletion resulting in adipocyte and osteoblast differentiation. This was shown to be due to *Gata2* inhibiting *Runx2* and β-catinin signalling thereby blocking osteoblastogenisis (X. Li *et al.*, 2016). Additionally, *Gata2* deletion in murine and human MSCs has been shown to result in an increase in G0/G1 cell cycle phase showing it to be involved in the regulation of proliferation (X. Li *et al.*, 2016). Mapping of the transcriptional landscape of murine BM at a single cell level has revealed considerable transcriptional remodelling of niche components occurs during stress conditions including adipocyte skewing of perivascular cells and down regulation of vascular Notch delta-like ligands resulting in a premature HSC induced myeloid transcriptional programme (Tikhonova *et al.*, 2019). As components of the niche play an essential role in regulating the inflammatory responses (Govindarajah and Reynaud, 2018; Mitroulis *et al.*, 2020), these changes are not represented in our current *Gata2*Vav murine model. These interactions could have a considerable impact during loss of *Gata2* providing important insights on disease progression in GATA2 immunodeficient patients. To overcome this limitation of our model, employing the germline *Gata2* mouse model (K.-W. W. Ling *et al.*, 2004; Rodrigues *et al.*, 2005) would be more representative of these germline mutations.

Finally, although we observed no differences in long or short-term HSCs during inflammation interestingly, it has been shown from previous unpublished data in our lab that age matched *Gata2*-het mice under steady-state conditions have substantial changes to the BM compartment (Abdelfattah, 2021). This data revealed a half reduction in *Gata2* heterozygous HSCs compared to control mice, which had significant increases in apoptosis and quiescence **(Figure S 5.22)** along with competitive BM transplantation from *Gata2* heterozygous HSCs, into primary recipients revealing a significant reduction in the multilineage repopulation *in* vivo and an inability to reconstitute secondary recipients during serial transplantation. This work in line with other studies illustrates *Gata2* heterozygous HSCs have a defective self-renewal capacity in line with previous work by (K.-W. W. Ling *et al.*, 2004; Rodrigues *et al.*, 2005).

When comparing this HSC data with the *Gata2*-VaviCre model during an inflammation stimulus outlined in this chapter, it is interesting to find that the deficit in HSCs observed in *Gata2*-het mice under steady-state, is not found during inflammation and these mice are able to produce comparable numbers of HSC as the pIpC treated control **(Figure 5.20).** In line with this, a similar observation can be made when comparing the Pre-GMP and GMP populations which once again are significantly reduced at steady state yet, are comparable to control mice during inflammation. This also holds true in our *in vitro* colony forming assay where we observed an equal number of CFU-GM production in *Gata2*-het BM cells compared with their reduction at steady state (**Figure 5.21)**. It is important to note that these are purely observations and would need to be validated by running these experiments side-by-side with further analysis. Nevertheless, this could suggest that *Gata2*-het HSCs and GM progenitors are inflammation-resistant-like cells.

The comparable levels of HSCs to pIpC treated controls observed in our *Gata2* mouse model along with the recovery of myeloid progenitors could prelude an expansion phase suggest that these HSC are activated during our chronic inflammatory setting and are skewed towards myeloid production as observed during emergency myelopoiesis (Takizawa, Boettcher and Manz, 2012; Clapes, Lefkopoulos and Trompouki, 2016b). This is supported by the observed decrease of $Gr1^{+}Mac1^{+}$ cells in the BM of which are increased in the peripheral blood **(Figures 5.9 and 5.19)**. This observation is in line with reports of HSC activation through IFN-γ during chronic *Myobacterium avium* infection or stimulation with LPS, resulting in increased proliferation due to cycling (Baldridge *et al.*, 2010). However, this is contrasting to findings by Pietras *et al.,* (2014) which highlights the length of IFN-stimulation via pIpC can have different effects on HSCs. In an acute setting this was shown to lead to a brief relaxation of quiescent enforcing mechanisms, resulting in transient proliferation, whilst a chronic stimulation promoted HSCs to re-enter quiescence, preventing the pool from exhaustion (Pietras, Warr and Passegué, 2011; Pietras *et al.*, 2014). This potential HSC proliferation could be specific to *Gata2* heterozygous HSCs and may indicate how GATA2 immunodeficient patients progress towards MDS or AML: sustained HSC activation during repetitive infections lead to the accumulation of genetic mutations and therefore disease progression. With this in mind, it would be interesting to see whether the proposed pIpC pulsing experiments over longer time periods in the Gata2 Vav mouse mode, would lead to disease initiation.

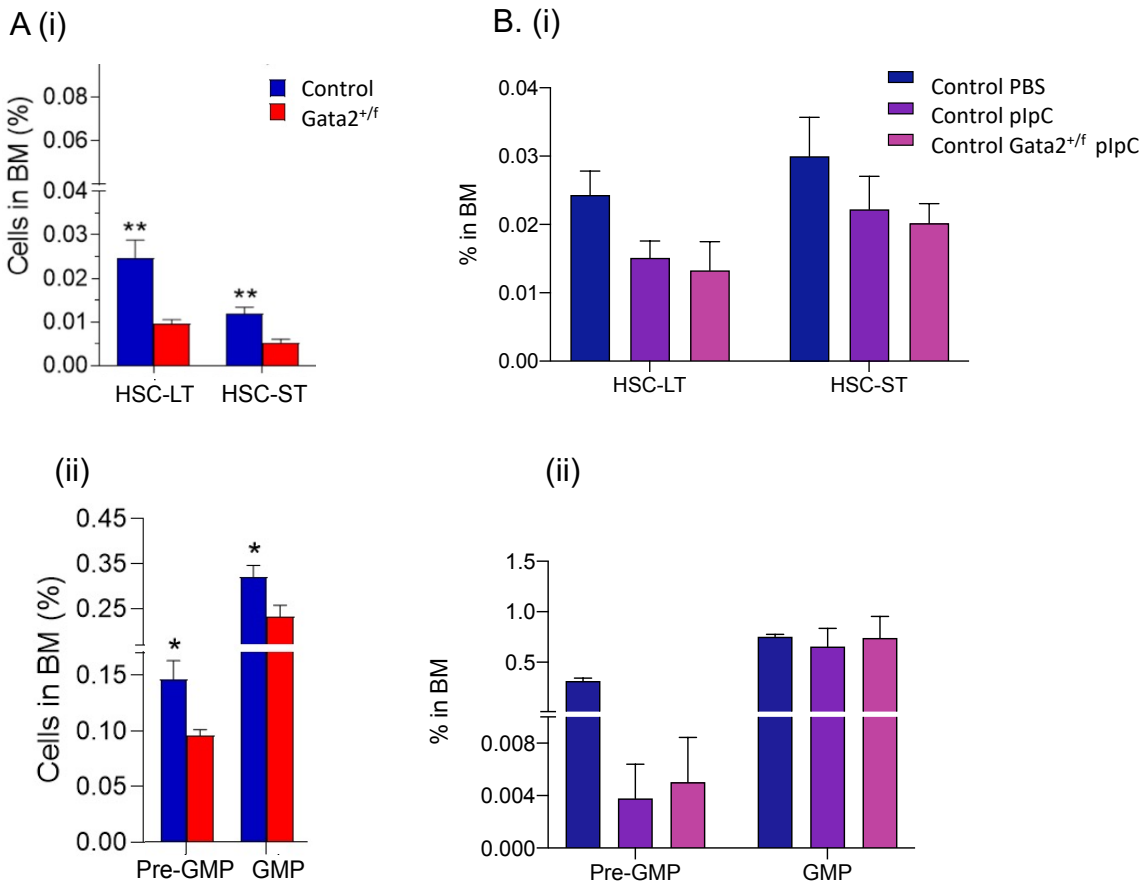


Figure 5.22: Modelling Gata2 haploinsufficiency using the *Gata2* Vav-iCe mouse model.

(A) Abundance of (i) long and short-term HSC (ii) pre-GMP and GMP cells within the BM of Gata2$^{+/fl}$ Vav-iCre and control mice at steady state (n=6). Data represent mean ± SEM. Statistical analysis: Mann-Whitney U test *, P < 0.05; **, P < 0.01. (Research conducted by Ali Abdelfattah) **(B)** Abundance of (i) long and short-term HSC and GMP cells within the BM of Gata2$^{+/fl}$ Vav-iCre and control mice treated with pIpC (or PBS) as shown (n=3,8,6).

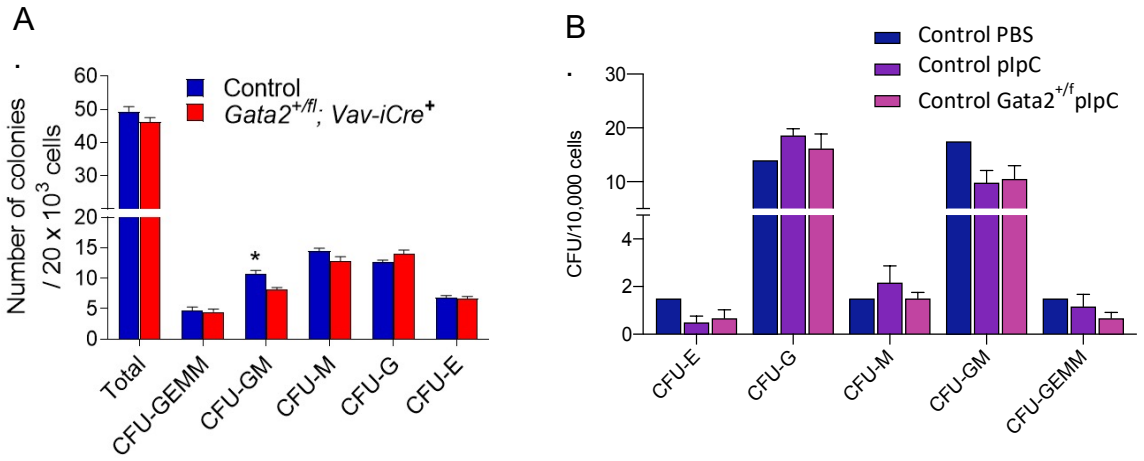


Figure 5.23: Modelling *Gata2* haploinsufficiency using the *Gata2* Vav-iCe mouse model. (A)CFC assays performed from total BM cells from Gata2$^{+/fl}$ Vav-iCre and control mice at steady state (n=4). Presented data are mean ± SEM. Statistical analysis: Mann-Whitney U test *, P < 0.05. (conducted by Ali Abdelfattah)**. (B)** CFC assays performed from total BM cells from Gata2$^{+/fl}$ Vav-iCre and control mice treated with pIpC (or PBS) as shown (n=1,4,4). Presented data are mean ± SEM. Statistical analysis: one-way ANOVA.

Chapter 6:

General Discussion

6.1 Towards a more specific model of GATA2 immunodeficiency

Unfortunately, over the course of this research we were unable to take our iPS model of GATA2 immunodeficiency and its progression towards MDS/AML to fruition. Nevertheless, iPSC technology remains an invaluable tool for modelling myeloid malignancy through; reprogramming of patient samples- providing a snapshot single somatic cells into stable cell lines or through the introduction of germline or somatic driver mutations into normal iPSCs- for the assessment of mutations in isolation or in combination with common secondary and tertiary molecular events.

This has been demonstrated in recent research conducted by the Papapetrou laboratory where multiple iPSCs have been generated from an AML patient containing genetic lesions of the specific AML; translocation and chromosome loss of 7q and KRAS mutation (G12D) (Chao *et al.*, 2017; Kotini *et al.*, 2017). During the reprogramming process the AML epigenome was reset, and leukaemia features re-established in HSPCs upon definitive-type *in vitro* differentiation where they exhibited, block in maturation, enhanced proliferation and survival and recapitulated molecular features of primary AML cells observed from transcriptome analysis. Strikingly, these AML-derived HSPCs were able to give rise to a transplantable leukaemia when engrafted into NSG mice (Kotini *et al.*, 2017; Wesely *et al.*, 2020). This study also highlights evidence of hierarchal organisation through the identification two populations; iLSC's and iBlasts, upon *in vitro* culture of HSPCs from AML derived iPSCs, supporting the LSC model proposed by Dick, (2008). These studies support how patient derived iPSC models of AML can closely recapitulate the LSC model and grant opportunities for the investigation of LSCs biological properties, whilst also allowing the identification and validation of specific AML therapeutic targets for clinical application, where other methods such as primary culture of AML samples are limited (Spyrou and Papapetrou, 2021). This again is supported by the proposed RUNX1-dependency in AML LSCs identified in these studies, that could hold potential for therapeutic targeting (Wesely *et al.*, 2020).

Showing what potentially could be achieved during our proposed stepwise model of GATA2 immunodeficiency the Papapetrou lab have also made great advances in how genetically engineered iPSC can be utilised to investigate clonal architecture and evolution in MDS and AML (Doulatov and Papapetrou, 2021). Their current work utilised iPSC engineering using CRISPR/Cas9 editing to develop a model of clonal haematopoiesis in order to capture distinct pre-malignant stages of clonal haematopoiesis (CH), followed by MDS and ultimately a transplantable leukaemia. Using ssRNA and ATAC sequencing techniques they elegantly demonstrated that these engineered iPSC lines upon haematopoietic differentiation, recapitulated transcriptional and chromatin accessibility signatures of primary human MDS

and AML, demonstrating the robustness of the model for identifying AML vulnerabilities. Moreover, the authors were able to characterise the transcriptional programmes driving progression to each disease stage and were able to identify dysregulation of inflammatory signalling as an early event which precedes leukemogenesis. Finally, further supporting the effectiveness of iPSC modelling for investigating therapeutic targets, this model identified an increased response of HSPCs to inflammatory ligands during LPS or IL-1β stimulation. This increased signalling was shown to be targetable using two preclinical compounds a dual inhibitor of IL-1 receptor-associated kinase 1 (*IRAK1*) and *IRAK4* and a small-molecule inhibitor of *UBE2N* (UC-764865), which inhibited haematopoietic progenitor colony formation from CH, MDS and AML stages (Doulatov and Papapetrou, 2021).

Thus far, two groups have attempted to utilise iPSCs to investigate how GATA2 mutations found in patients alter haematopoietic development. Kotini *et al.,* (2017) was able to produce a stepwise model of Gata2 immunodeficiency, towards a low and high risk MDS stage. Cells from a patient with high risk MDS were reprogrammed, capturing a clone with an isolated GATA2 T357N germline mutation and a second clone representing a more evolved disease stage with three additional driver mutations. (Kotini *et al.*, 2017). The second study conducted by Jung *et al.,* (2018) successfully reprogrammed cells from patients harbouring two missense mutations: R337X and R361H along with a mutation in intron 5 resulting in aberrant splicing. During both studies very subtle defects were observed during haematopoietic commitment with modest reductions in clonogenicity observed in CFC assays, specifically in BFU-E, CFU-GEMM. Nevertheless, the characteristic defects in monocytes and NK cells observed in GATA2 patients were not recapitulated, although investigation into B and T lineages were not carried out. The lack of phenotype presented in these studies could be attributed to the lack of potential co-operating factors that are potentially important for mutant GATA2 protein in initiating the hallmark BM failure observed in patients such as immunological challenge from viral and bacterial pathogens and

The current protocols used for the differentiation of iPSCs into the haematopoietic lineage follow a linear progression of mesoderm commitment to embryonal HSPCs and therefore these models exhibit a degree of developmental immaturity (Jung *et al.*, 2018; Wang *et al.*, 2021a). Additionally, the assessment of self-renewal and multilineage engraftment potential in these models are beyond the current scope of iPSC technology due to inability to produce engraftable HSPC without enforced expression of various TF (Jung, Dunbar and Winkler, 2015; Sugimura *et al.*, 2017). Furthermore, many GATA2 patients experience long latency before cytopenias develop and disease presentation occurs. In support of this notion, although NK are largest abolished in the BM of GATA2 patients, residual NK cells can be found in the

PB which decline over time and during disease progression (Bigley *et al.*, 2011; Dickinson *et al.*, 2014; Spinner *et al.*, 2014). Unfortunately, iPSC models would be unable to capture this progressive loss of certain lineages which can take many years to develop due the short time frame in which iPSCs are able to be directed into different haematopoietic lineages. Finally generating iPSCs from to capture specific stages of disease from GATA2 patients aren't without challenges. Many genetic aberrations are not compatible with the reprogramming process as observed from these studies where the authors had difficulty in reprogramming cells that captured *STAG2* secondary mutations and chromosome 7q. To overcome this limitation lines were engineered by deleting chromosome 7q in GATA2 iPSC (Kotini *et al.*, 2017; Jung *et al.*, 2018).

Loss of function techniques such as RNAi-mediated knockdown, CRISPR/Cas9 mediated Knock-out and transgenic mouse models have been used for many years to investigate physiological significance of disease relevant mutations. However, though GATA2 mutations that cause the immunodeficiency syndrome are LOF, studies have shown that some mutations are dominant-negative mutations (Shimizu and Yamamoto, 2020; Oleaga-Quintas *et al.*, 2021) and therefore these complete functional knock-out models may mask potential phenotypes. This is an additional limitation of methodology employed in chapter 5. To overcome these limitations, knock-in murine models can be generated using the CRISPR/Cas9 system to introduce specific disease-causing gene variants to more faithfully recapitulate pathogenic potential. However, it is important to note that some pathogenic mutations have been shown to present distinct species-specific phenotypes such as the Schlafen 14 (SLFN14) mutation which leads to an inherited macrothrombocytopenia and bleeding disorder in human patients, which showed no changes to platelet or erythroid function but presented animals with severe microcytic erythrocytosis (Hyman and Tanzi, 2019; Stapley *et al.*, 2021). These more specific murine models allow the crosstalk of cells with their niche components which could be potentially important for modelling GATA2 immunodeficiency and its progression.

It has been indicated that timing of mutation acquisition along with patterns of mutation co-occurrence are not random and yet are currently understudied in the literature (Doulatov and Papapetrou, 2021). In regard to GATA2 immunodeficiency there is still some controversy over which genetic events preclude each other, in particular acquisition of ASXL1 heterozygous mutations and loss of chromosome 7 (Bödör *et al.*, 2012; West *et al.*, 2014; Wlodarski, Collin and Horwitz, 2017). iPSC engineered lines with these specific genetic events in different combination would be able to shed light on this issue. Additionally, to more closely resemble the stepwise acquisition, these secondary events can be engineered with inducible control using cre-mediated recombination or tetracycline control (Stokes, Dominguez-Sola and

Papapetrou, 2018; Doulatov and Papapetrou, 2021). Though large-scale genetic lesions are challenging to engineer Kotini *et al,* have reported two approaches using a modified cre-loxp engineering strategy utilising two inverted loxP sites in conjunction with an HSV-tk transgene into chromosome 7q or alternatively a CRISPR editing approach using dual CRISPR-mediated DNA breaks (Kotini *et al.*, 2015; Kotini and Papapetrou, 2020). Alternatively, sequentially inducible mouse models incorporating these specific mutations can also be employed in order to model longer latency between mutations, which might be necessary for disease progression that cannot be achieved in human iPSC models as outlined previously. This has been outlined in Dnmt3a and nucleophosmin 1 (Npm1) inducible mutations in a murine model of human CH and AML in which CH progresses to myeloproliferative disorder (MPD) upon induction of *Npm1* and was shown to drive a more aggressive MPD with longer latency, which during simultaneous expression of these mutations failed to cause lethal malignancy (Guryanova *et al.*, 2016; Loberg *et al.*, 2019).

Given the strengths and weaknesses of each modelling technique it becomes quite apparent that scientific investigation of GATA2 syndromes would benefit from combined investigation using murine and human models employing specific GATA2 mutations of which limitations of one method can be overcome by another. For example, current limitations of modelling haematopoietic cells and their interactions with the niche components during iPSC modelling and primary patient culture makes mouse models carrying specific patient mutations a more reliable method for interpreting biological implications of inflammation during these syndromes. Whilst other aspects of the disease are better suited to human models such as investigating cell intrinsic features of inflammation during engineered iPSC models in conjunction with patient samples for more robust conclusions, along with developing therapeutic targets.

6.2 Inflammation is an important factor in GATA2 syndrome disease progression

The importance of inflammation in disease initiation and progression in immunological deficiency syndromes, CHIP and haematological malignancies such MDS and AML have gained much traction over the last few years (Sallman and List, 2019; Cook, Luo and Rauh, 2020; Matos, Magalhães and Rauh, 2020). In the final chapter of this thesis, we present how chronic inflammation over just one week, causes dramatic changes to the phenotype of *Gata2* heterozygous mice, with prospective inflammation resistant populations such as HSCs warranting further investigation, over longer time periods with additional analysis. These dramatic changes highlight how inflammation during the immunodeficiency stage of GATA2 patients cannot be ignored as a biological implication during disease initiation. In agreement with this preliminary observation, resistance to the damaging effects of inflammation has been

a common theme specifically observed in *TET2* and *DNMT3a* clonal haematopoiesis (Cull *et al.*, 2017b; Abegunde *et al.*, 2018; Cai *et al.*, 2018). In a recent publication by the King Lab, *Batf2* and *Dnmt3a* deficient stem cells where shown to resist depletion upon chronic infection, during secondary stress induction in a post transplantation mosaic mouse model (Hormaechea-Agulla *et al.*, 2021). It is important to note that the phenotype observed in this model was revealed through inducing rapid rates of cell division which exaggerates competitive differences from the mosaic mice, that would take a long latency in non-transplant setting. Once again this is very interesting as the resistant populations observed in our *in* vivo model is profound considering the relatively low stimulation applied to these mice in a non-transplant setting.

Another factor that needs to be taken into consideration, is whether the type of infectious stimuli GATA2 patients encounter affects how *Gata2* mutant cells respond to inflammation. There have been a broad spectrum of viral, bacterial and fungal infections that have been documented in patients with GATA2 syndromes (**Table 1.12**) (Dickinson *et al.*, 2014; Spinner *et al.*, 2014; Collin, Dickinson and Bigley, 2015b) and it is very well established that immune responses to these insults activate different immune signalling pathways (Alexopoulou *et al.*, 2001; Kawai and Akira, 2008; Peignier and Parker, 2021). Therefore, investigation of different infection types such as mycobacteria strains- implicated in the GATA2 MonoMac syndrome, viral and fungal infections may have very distinct effects on disease outcome, which could further differ with the different GATA2 mutations modelled due to the dominant negative effects some mutations have been demonstrated to have (Chong *et al.*, 2018). This idea of differential response to stressors from mutations has observed in *PPN1D* mutant clones in CH. Hsu, *et al* 2018 presented that *PPN1D* mutant cells in the absence of environmental stress confer no selective advantage over their wild-type counterparts, nevertheless, following treatment with Cisplatin these cells are more resistance to apoptosis, resulting in a selective advantage over multiple rounds of exposure. Importantly, this advantage was observed with specific chemotherapeutic agents such Cisplatin and Doxorubicin but not Vincristine, demonstrating the risk of CH progression in patients carrying *PPN1D* mutations with particular chemotherapeutic agents (Hsu *et al.*, 2018). In further support of this idea Hormaechea-Agulla *et al.*, 2021 have reciently presented how the expansion of *Dnmt3a*$^{-/-}$ HSCs occurred specifically during stimulation with *M.avium* in comparison to acute lymphochoriomeningitis virus (Armstrong strain), LPS, TNF-α, or pIpC in mice, revealing that sustained IFN-γ was essential for this expansion. Identifying underlying molecular mechanisms of GATA2 mutant cells during specific environmental insults could have important prognostic value for GATA2 mutant patients in predicting which infections may result in poorer disease outcomes and could identify potential biomarkers that can aid clinicians with early diagnosis and treatment

intervention at these earlier stages of immunodeficiency- which currently are a limiting factor in GATA2 patient identification and management.

6.3 Molecular and biological implications for GATA2 immunodeficiency- What do we know?

Whilst there is a lack of molecular information in the literature about how GATA2 mutations lead to the clinical heterogeneity found in patients, a recent paper employing whole transcriptome sequencing of HSPCs from eight patients harbouring pathogenic GATA2 mutations in an MDS stage of disease has provided key insight. In line with heterozygous mouse models of Gata2 that have shown that HSCs have increased quiescence, are more susceptible to apoptosis and have reduced repopulating capacity (K. W. Ling *et al.*, 2004; Rodrigues *et al.*, 2005; Rodrigues, Boyd, Fugazza, May, Yan Ping Guo, *et al.*, 2008), Wu *et al.,* 2020 highlighted defects in gene pathways that control the exit from quiescence such as targets of *E2F* and *MYC* and genes involved in apoptosis. Furthermore, defects in genes implicated in DNA replication and repair, cell cycle were also reported. Many case studies have reported the loss of LMPP and a significant reduction of GMP compartments in GATA2 patient immunophenotyping which has been somewhat recapitulated in murine studies (K. W. Ling *et al.*, 2004; Rodrigues *et al.*, 2005; Rodrigues, Boyd, Fugazza, May, Yan Ping Guo, *et al.*, 2008; Bigley *et al.*, 2011; Spinner *et al.*, 2014). This finding was thought to be surprising given the relatively low expression of GATA2 in these cells. However, work by Wu *et al* 2020 revealed myeloid/lymphoid genes to be reduced in patient HSPCs whilst erythoid/megakaryocyte programs were relatively well preserved which has been further supported by ssRNA sequencing of GATA2 mutant iPSC cells (Jung *et al.*, 2018). This reduction was suggested to be mediated through the downregulation of Spi-1 Proto-Oncogene (SPI1) in HSCs and GMPs observed in GATA2 patients (Wu *et al.*, 2019). *SPI1* is the gene encoding PU.1 which plays an antagonistic role with GATA2 during lineage commitment where PU.1 directs HSCs to the lymphoid myeloid progenitors (LMPP) whilst interacting with GATA1 to inhibit commitment to megakaryocyte erythroid progenitor (MEP) (Scott *et al.*, 1994; Friedman, 2007; Fiedler and Brunner, 2012). As GATA2 regulates transcription through binding to gene promoters or alternatively interacting with other proteins in various complexes in a context dependent manner, this could explain how reduced GATA2 dosage leads to aberrant protein partner interactions leading to the lineage skewing observed in patients. This is further supported by the observation that *SPI1* was also found to be down regulated in $CD34^+$ and mononuclear cells in patient PB (Bigley *et al.*, 2011; Cortes-Lavaud *et al.*, 2015). The transition from CMP to GMP is regulated by PU.1 and CEBPα levels, along with the lineage choice between monocytes and granulocytes. Many functional studies have observed that reduced expression of PU.1 results in increased granulopoiesis at the expense of

monocyte production, along with gene expression analysis in PU.1-deficient progenitors revealing reduced or absent expression of several monocyte-specific genes such as M-CSF (Fiedler and Brunner, 2012). The transcription changes outlined by Wu *et al.*, 2020 could also explain the loss of monocytes and preservation of granulocytes observed in GATA2 patients and warrants more in-depth investigations.

Collectively, this study highlights that GATA2 mutations found in patients cause global dysregulation of transcriptional programs throughout the haematopoietic system. Given the differential epigenetic regulation of the GATA2 gene highlighted by differing methylation patterns at the GATA2 promotor region between patients and the abarrent interactions with protein partners such as PU.1 in dominant negative mutant proteins (Chong *et al.*, 2018; Galera *et al.*, 2018; Kim *et al.*, 2019), future studies employing transcriptional profiling at the single cell level, in bigger cohorts would be important to assess how these gene programs differ between these parameters. This ultimately could account in part, for the vast clinical heterogeneity observed in the clinic.

6.4 Towards better clinical management of GATA2 syndromes

The heterogenous presentation of GATA2 deficiency severely complicates medical management of these patients. In particular, the majority of these patients present with no clinical manifestation before the age of 20, nevertheless it is important to highlight that paediatric cases do present with infections such as HPV and mycobacteria and MDS or AML (Hsu *et al.*, 2011; Luesink *et al.*, 2012; Spinner *et al.*, 2014; Wang *et al.*, 2015). Therefore, early identification of GATA2 syndromes can be challenging and is heavily reliant on clinical specialists being educated on the key events that occur prior to phenotype presentation. One of the earliest findings in GATA2 patients is the onset of cytopenia, in particular neutropenia (with or without a history of mild or intermittent infections) and monocytopenia are hallmark primary features which can be identified from routine CBC analysis (Bigley *et al.*, 2011; Spinner *et al.*, 2014; Collin, Dickinson and Bigley, 2015a). Careful interpretation of these CBCs is invaluable for highlighting potential suspicious parameters caused by an underlying GATA2 mutation such as the detection of macrocytic anaemia with normal folic acid and B12 levels (Bödör *et al.*, 2012; Al Seraihi *et al.*, 2018; Donadieu *et al.*, 2018). Granted that decreases in WBC can be a mark any number of conditions with derived from familial syndromes or underlying errors of immunity which can complicate diagnosis (Saettini *et al.*, 2020), leukopenia with additional monocytopenia should prompt screening for GATA2 as an underlying cause (Saettini *et al.*, 2020). In contrast many familial germline predispositions such as those caused by *RUNX1*, *ANKRD26*, *SAMD9*, *SAMD9L* and *ETV6* carry the hallmark of thrombocytopenia of which in GATA2 patients occurs at much later stages (Owen *et al.*,

2008; Bluteau *et al.*, 2014; Bannon and Dinardo, 2016; Melazzini *et al.*, 2016; Feurstein and Godley, 2017).

Treatment intervention at the immunodeficiency stage prior to the progression towards myeloid neoplasm occurring in 75% of patients is highly desirable (van Lier *et al.*, 2020). However, as with other immunodeficiency disorder this remains challenging as it requires careful balancing of immunosuppression in individuals with increased risk of infection (Delmonte *et al.*, 2019). As we have presented preliminary evidence supporting the idea that loss of *GATA2* in the haematopoietic compartment confers a differential response to inflammation, this could suggest that GATA2 patients may benefit from therapeutic strategies that control hyperactivation of the immune system. Many independent mechanisms have been outlined to contribute to hyperactivation of immune signalling pathways observed in MDS and AML (Barreyro, Chlon and Starczynowski, 2018b; Sallman and List, 2019). These have been highlighted to converge on the central complex involving *IRAK1*, *IRAK4* and *TRAF6* (Fang *et al.*, 2017; Smith *et al.*, 2019; Muto *et al.*, 2020). Therefore, inhibition of this complex has great therapeutic potential and this strategy has been demonstrated through targeting using various inhibitors or antibodies in pre-clinical studies (Hosseini *et al.*, 2018; Mitchell *et al.*, 2018; Smith *et al.*, 2019; Wang *et al.*, 2021b). Clinical testing of *IRAK1* and *IRAK4* inhibitors are currently underway for the treatment of MDS, MPN, AML, lymphoma along with immune and inflammatory disorders of which GATA2 patients could potentially benefit, once pathophysiology at the molecular level has been established.

The only curative treatment currently available for GATA2 deficiency patients is allogenic haematopoietic stem cell transplant (HSCT). Nevertheless, clinical data focused on parameters such as optimal donor type, stem cell source and conditioning regimes are yet to define an optimal strategy (Ramzan *et al.*, 2017; Bogaert *et al.*, 2020; van Lier *et al.*, 2020). Furthermore, there has also been much conversation focused around whether earlier transplantation in young adults with low risk MDS and germline GATA2 mutations would be more beneficial (Cutler *et al.*, 2004; Steensma, 2019). This is a complex issue of which many factors require careful consideration, such as degree of immunocompromise or recurrent infections in each individual case. Thorough evaluation of infectious and inflammatory comorbidities pre transplant are vital as viral pathogens during HSCT carry a considerable risk for morbidity and mortality in recipients. This is largely subject to immunosuppressive therapy individuals receive prior to transplantation accompanied by the delay in adaptive immune system reconstitution post-transplantation (van Lier *et al.*, 2020; Annaloro *et al.*, 2021). Whilst Cellular immunodeficiency is a parameter which is closely monitored and is used as an indicator for when to initiate HSCT, many patients remain stable for years before these

parameters crash. Moreover, HSCT at earlier stages of disease can have detrimental impacts later on in life such as acquisition of therapy related AML (t-AML), a complication that arises after exposure to chemotherapy and or radiation which accounts for approximately 10% of all AMLs considered to have a poorer out-come (Schoch *et al.*, 2004; Borthakur and Estey, 2007; Godley and Larson, 2008; Kayser *et al.*, 2011). Another important factor that needs to be taken into account recently highlighted in the literature, is the risk of donors with synonymous *GATA2* variants being taken forward as potential HSCT donors (Galera *et al.*, 2018). Synonymous mutations have been largely reported in the literature as non-pathogenic due to the lack of amino acid change and account for 17% of all reported variants in GATA2 mutations (Shimizu and Yamamoto, 2020). However, cases have recently been observed to cause immunodeficiency, severe infections and lung disease due to RNA-deleterious effects such as splicing errors leading to RNA and protein reduction and selective loss of mutant RNA (Wehr *et al.*, 2018; Kozyra *et al.*, 2020). To ensure that these patients are not overlooked during genetic counselling it has been proposed that following inconclusive DNA analysis in patients with hallmark features of GATA2 deficiency diagnostic sequencing should be integrated with additional RNA sequencing to identify these potential variants (Kozyra *et al.*, 2020).

Taken together, better patient outcome in the future, is heavily reliant on the dual efforts of scientists and clinicians. There is an unmet need for scientists to build more specific and biologically relevant models of GATA2 immunodeficiency and its progression to haematological malignancies. These are invaluable for gaining important insight into pathophysiology using molecular tools such as scRNA and ATAC sequencing, which can then be exploited as potential diagnostic markers or for investigating potential therapeutic targets. In parallel, educating clinicians on key events that occur prior to phenotype presentation or when intermittent infections known to be associated with GATA2 syndromes are implicated (such as mycobacterial infections highlighted in MonoMac syndrome) are essential for improved timing of early referral to genetic counselling. Finally, comprehensive screening of patients and family members are fundamental in prevent HSCT from donors with pathogenic or synonymous variants. Additionally, to advance research investigating GATA2 immunodeficiency syndromes and its progression, the research community could benefit from a GATA2 research foundation akin to the RUNX1 Research program which was founded in 2016. This foundation was set up to raise awareness of patients with RUNX1-FPD through funding world class collaborative research. It has partnered with various charities such as Leukaemia and Lymphoma Society (LLS) providing grants of $200,000 per year for translational leukaemia research related to this familial RUNX1 disorder along with funding for

new and innovative research into therapies to target the RUNX1 disorder before disease progression is reached (https://www.runx1-fpd.org/research).

The main changes that I would have incorporated into my experimental approach with the benefit of hindsight are:

iPSC GATA2 modelling:

- A second round of CRISPR/Cas9 editing in iPSC clone 62 or 66 (Figure 4.7 A) containing one GATA2 allele with the T354M mutation:
    - Design the guide RNA to target the second GATA2 allele modified by NHEJ.
    - Deliver a HDR template containing a wild-type sequence to repair the disrupted GATA2 allele producing a human iPSC GATA2 $^{+/T354M}$ cell line.
- Use an inducible approach for secondary ASXL1 mutations such as cre-mediated recombination or tetracycline control.
    - To more faithfully recapitulate the germline GATA2 mutations followed by acquisition of somatic secondary ASXL1 mutations in patients.

Inflammation during shRNA GATA2 knockdown in CD34 HSPCs

- Treatment of cells with IFN-1 to represent viral challenge alongside LPS experiments (mirroring bacterial challenge) to assess whether infection type has a differential effect.

Inflammation during Gata2 VaviCre model

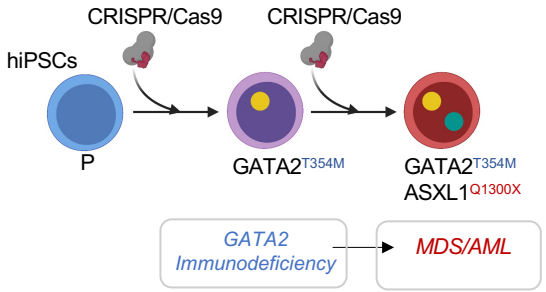
- Incorporated the mouse control group: Gata2$^{+/f}$ Vav positive animals treated with PBS.
- Pulse treatment of animals with pIpC to represent recurrent infections
    - Long term analysis of one experimental cohort to assess if animals develop BM failure or disease progression
    - Short term analysis of a second experimental cohort: immunophenotyping to assess changes to haematopoiesis in the BM and spleen.
    - Additional CBC analysis to assess haemoglobin (Hgb) and haematocrit (Hct) values to measure the proportion of RBCs which were excluded in the analysis through gating strategies in the spleen and BM and RBC lysis in PB analysis.
- Stimulate animals with *Mycobacterium* infection along-side pIpC treatment to see the effect of different infection type in disease progression.

Graphical Abstract

1 **Modelling** GATA2 immunodeficiency syndromes and its progression towards MDS and AML.

Approach:
CRISPR Cas9 and base editing of human iPSCs to engineer specific iPSC lines at immunodeficiency and MDS/AML stages of disease.

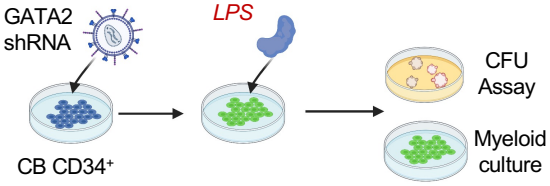
Further experiments required
- Second round of CRISPR editing to repair second disrupted GATA2 allele.



2 **Hypothesis**: Inflammation causes a selection pressure during GATA2 immunodeficiency for disease progression.

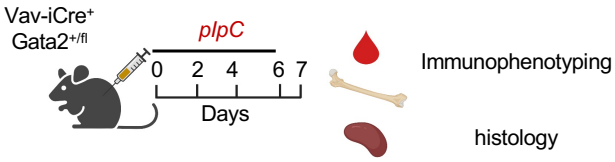
Approach 1:
LPS stimulation in CB CD34⁺ HSPCs mirroring bacterial infections in patients

Key Findings
- Increased BFU-E output in CFC assays
- Reduced Go and increase CD14 expression in myeloid cultures



Approach 2:
pIpC stimulation in GATA2 heterozygous mice mirroring viral infections in patients

Key Findings:
- Splenomegaly and granuloma formation
- Decreased myeloid cells in BM
- HSPCs adapt to inflammation



References

1. Abdel-Wahab, O. *et al.* (2012) 'ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression', *Cancer Cell*. Elsevier, 22(2), pp. 180–193. doi: 10.1016/j.ccr.2012.06.032.
2. Abegunde, S. O. *et al.* (2018) 'An inflammatory environment containing TNFα favors Tet2-mutant clonal hematopoiesis', *Experimental Hematology*. Elsevier Inc., 59, pp. 60–65. doi: 10.1016/j.exphem.2017.11.002.
3. Acar, M. *et al.* (2015) 'Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal', *Nature*. Nature Publishing Group, 526(7571), pp. 126–130. doi: 10.1038/nature15250.
4. Ackermann, M. *et al.* (2015) 'Lost in translation: pluripotent stem cell-derived hematopoiesis.', *EMBO molecular medicine*, 7(11), pp. 1388–402. doi: 10.15252/emmm.201505301.
5. Adewumi, O. *et al.* (2007) 'Characterization of human embryonic stem cell lines by the International Stem Cell Initiative', *Nature Biotechnology*. Nature Publishing Group, 25(7), pp. 803–816. doi: 10.1038/nbt1318.
6. Adli, M. (2018) 'The CRISPR tool kit for genome editing and beyond', *Nature Communications*. Nature Publishing Group, pp. 1–13. doi: 10.1038/s41467-018-04252-2.
7. Adolfsson, J. *et al.* (2005) 'Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: A revised road map for adult blood lineage commitment', *Cell*. Cell Press, 121(2), pp. 295–306. doi: 10.1016/j.cell.2005.02.013.
8. Akbar, M. A. *et al.* (2016) 'ARC Syndrome-Linked Vps33B Protein Is Required for Inflammatory Endosomal Maturation and Signal Termination', *Immunity*. Cell Press, 45(2), pp. 267–279. doi: 10.1016/j.immuni.2016.07.010.
9. Albiero, M. *et al.* (2015) 'Bone marrow macrophages contribute to diabetic stem cell mobilopathy by producing oncostatin M', *Diabetes*. American Diabetes Association Inc., 64(8), pp. 2957–2968. doi: 10.2337/db14-1473.
10. Alexopoulou, L. *et al.* (2001) 'Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3', *Nature*. Nature Publishing Group, 413(6857), pp. 732–738. doi: 10.1038/35099560.
11. Alfayez, M. *et al.* (2020) 'CPX-351 (vyxeos) in AML', *Leukemia and Lymphoma*. Taylor and Francis Ltd, pp. 288–297. doi: 10.1080/10428194.2019.1660970.
12. Ali, F. *et al.* (2019) 'Plasmalogens Inhibit Endocytosis of Toll-like Receptor 4 to Attenuate the Inflammatory Signal in Microglial Cells', *Molecular Neurobiology*. Humana Press Inc., 56(5), pp. 3404–3419. doi: 10.1007/s12035-018-1307-2.
13. Allen, F. *et al.* (2019) 'Predicting the mutations generated by repair of Cas9-induced double-strand breaks', *Nature Biotechnology*. Nature Publishing Group, 37(1), pp. 64–82. doi: 10.1038/nbt.4317.
14. Alsayegh, K. *et al.* (2019) 'Hematopoietic Differentiation of Human Pluripotent Stem Cells: HOX and GATA Transcription Factors as Master Regulators', *Current Genomics*. Bentham Science Publishers Ltd., 20(6), pp. 438–452. doi: 10.2174/1389202920666191017163837.
15. Alvarez Argote, J. and Dasanu, C. A. (2018) 'ASXL1 mutations in myeloid neoplasms: pathogenetic considerations, impact on clinical outcomes and survival', *Current Medical Research and Opinion*. Taylor and Francis Ltd, pp. 757–763. doi: 10.1080/03007995.2016.1276896.
16. Annaloro, C. *et al.* (2021) 'Viral Infections in HSCT: Detection, Monitoring, Clinical Management, and Immunologic Implications', *Frontiers in Immunology*. Frontiers Media S.A., p. 3482. doi: 10.3389/fimmu.2020.569381.
17. Antony-Debré, I. *et al.* (2012) 'MYH10 protein expression in platelets as a biomarker of RUNX1 and FLI1 alterations', *Blood*. American Society of Hematology, 120(13), pp. 2719–2722. doi: 10.1182/blood-2012-04-422352.
18. Antony-Debré, I. *et al.* (2015) 'Level of RUNX1 activity is critical for leukemic predisposition but not for thrombocytopenia', *Blood*. American Society of Hematology, 125(6), pp. 930–940. doi: 10.1182/blood-2014-06-585513.
19. Anzalone, A. V., Koblan, L. W. and Liu, D. R. (2020) 'Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors', *Nature Biotechnology*. Nature Research, pp. 824–844. doi: 10.1038/s41587-020-0561-9.
20. Appelbaum, F. R. and Bernstein, I. D. (2017) 'Gemtuzumab ozogamicin for acute myeloid leukemia', *Blood*. American Society of Hematology, pp. 2373–2376. doi: 10.1182/blood-2017-09-797712.
21. Arber, D. A. *et al.* (2016) 'The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia', *Blood*. American Society of Hematology, pp. 2391–2405. doi:

10.1182/blood-2016-03-643544.
22. Asada, S. *et al.* (2018) 'Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis', *Nature Communications*. Nature Publishing Group, 9(1), p. 2733. doi: 10.1038/s41467-018-05085-9.
23. Avellino, R. *et al.* (2016) 'An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation', *Blood*. The American Society of Hematology, 127(24), pp. 2991–3003. doi: 10.1182/blood-2016-01-695759.
24. Avellino, R. and Delwel, R. (2017) 'Expression and regulation of C/EBPα in normal myelopoiesis and in malignant transformation.', *Blood*. American Society of Hematology, 129(15), pp. 2083–2091. doi: 10.1182/blood-2016-09-687822.
25. Avior, Y., Sagi, I. and Benvenisty, N. (2016) 'Pluripotent stem cells in disease modelling and drug discovery', *Nature Reviews Molecular Cell Biology*, pp. 170–182. doi: 10.1038/nrm.2015.27.
26. Bai, H. *et al.* (2017) 'Cytoplasmic transport and nuclear import of plasmid DNA', *Bioscience Reports*. Portland Press Ltd. doi: 10.1042/BSR20160616.
27. Baker, D. E. C. *et al.* (2007) 'Adaptation to culture of human embryonic stem cells and oncogenesis in vivo', *Nature Biotechnology*. Nature Publishing Group, pp. 207–215. doi: 10.1038/nbt1285.
28. Balasubramani, A. *et al.* (2015) 'Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex', *Nature Communications*. Nature Publishing Group, 6(1), pp. 1–15. doi: 10.1038/ncomms8307.
29. Baldridge, M. T. *et al.* (2010) 'Quiescent haematopoietic stem cells are activated by IFN-γ in response to chronic infection', *Nature*. Nature, 465(7299), pp. 793–797. doi: 10.1038/nature09135.
30. Balkwill, F., Charles, K. A. and Mantovani, A. (2005) 'Smoldering and polarized inflammation in the initiation and promotion of malignant disease', *Cancer Cell*. Cell Press, pp. 211–217. doi: 10.1016/j.ccr.2005.02.013.
31. Balkwill, F. and Mantovani, A. (2001) 'Inflammation and cancer: Back to Virchow?', *Lancet*. Elsevier Limited, pp. 539–545. doi: 10.1016/S0140-6736(00)04046-0.
32. Bannon, S. A. and Dinardo, C. D. (2016) 'Hereditary Predispositions to Myelodysplastic Syndrome', *Int. J. Mol. Sci*, 17. doi: 10.3390/ijms17060838.
33. Barreyro, L., Chlon, T. M. and Starczynowski, D. T. (2018a) 'Chronic immune response dysregulation in MDS pathogenesis', *Blood*. doi: 10.1182/blood-2018-03-784116.
34. Barreyro, L., Chlon, T. M. and Starczynowski, D. T. (2018b) 'Chronic immune response dysregulation in MDS pathogenesis', *Blood*. American Society of Hematology, 132(15), pp. 1553–1560. doi: 10.1182/blood-2018-03-784116.
35. Bassett, A. R. (2017) 'Editing the genome of hiPSC with CRISPR/Cas9: disease models', *Mammalian Genome*, pp. 1–17. doi: 10.1007/s00335-017-9684-9.
36. Bates, D. L. *et al.* (2008) 'Crystal Structures of Multiple GATA Zinc Fingers Bound to DNA Reveal New Insights into DNA Recognition and Self-Association by GATA', *Journal of Molecular Biology*. Academic Press, 381(5), pp. 1292–1306. doi: 10.1016/j.jmb.2008.06.072.
37. Bazeley, P., Morales, R. and Tang, W. H. W. (2020) 'Evidence of Clonal Hematopoiesis and Risk of Heart Failure', *Current Heart Failure Reports*. Springer, pp. 271–276. doi: 10.1007/s11897-020-00476-w.
38. Becker, A. J., McCulloch, E. A. and Till, J. E. (1963) 'Cytological demonstration of the clonal nature of spleen colonies derived from transplated mouse marrow cells', *Nature*, 197, pp. 452–454.
39. Bejar, R. *et al.* (2012) 'Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 30(27), pp. 3376–3382. doi: 10.1200/JCO.2011.40.7379.
40. Bellissimo, D. C. *et al.* (2020) 'Runx1 negatively regulates inflammatory cytokine production by neutrophils in response to Toll-like receptor signaling', *Blood Advances*. American Society of Hematology, 4(6), pp. 1145–1158. doi: 10.1182/bloodadvances.2019000785.
41. Bellissimo, D. C. and Speck, N. A. (2017) 'RUNX1 mutations in inherited and sporadic leukemia', *Frontiers in Cell and Developmental Biology*, 5(111), pp. 1–11. doi: 10.3389/fcell.2017.00111.
42. Ben-Ami, O. *et al.* (2013) 'Addiction of t(8;21) and inv(16) Acute Myeloid Leukemia to Native RUNX1', *Cell Reports*, 4(6), pp. 1131–1143. doi: 10.1016/j.celrep.2013.08.020.
43. Benveniste, P. *et al.* (2010) 'Intermediate-Term Hematopoietic Stem Cells with Extended but Time-Limited Reconstitution Potential', *Cell Stem Cell*. Cell Press, 6(1), pp. 48–58. doi: 10.1016/j.stem.2009.11.014.
44. Bereshchenko, O. *et al.* (2009) 'Hematopoietic Stem Cell Expansion Precedes the Generation of Committed Myeloid Leukemia-Initiating Cells in C/EBPα Mutant AML', *Cancer Cell*. Elsevier, 16(5), pp. 390–400. doi: 10.1016/j.ccr.2009.09.036.

45. Bernasconi, P. *et al.* (2005) 'Incidence and prognostic significance of karyotype abnormalities in de novo primary myelodysplastic syndromes: A study on 331 patients from a single institution', *Leukemia*. Nature Publishing Group, 19(8), pp. 1424–1431. doi: 10.1038/sj.leu.2403806.
46. Béziat, V. *et al.* (2010) 'NK cell terminal differentiation: Correlated stepwise decrease of NKG2A and acquisition of KIRs', *PLoS ONE*. Public Library of Science, 5(8). doi: 10.1371/journal.pone.0011966.
47. Bhatia, M. *et al.* (1997) 'Quantitative analysis reveals expansion of human hematopoietic repopulating cells after short-term ex vivo culture.', *The Journal of experimental medicine*. Rockefeller University Press, 186(4), pp. 619–24. doi: 10.1084/jem.186.4.619.
48. Bigley, V. *et al.* (2011) 'The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency', *Journal of Experimental Medicine*. The Rockefeller University Press, 208(2), pp. 227–234. doi: 10.1084/jem.20101459.
49. Bluteau, D. *et al.* (2014) 'Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 124(2), pp. 580–591. doi: 10.1172/JCI71861.
50. Bluteau, O. *et al.* (2017) 'A landscape of germline mutations in a cohort of inherited bone marrow failure patients.', *Blood*. American Society of Hematology, p. blood-2017-09-806489. doi: 10.1182/blood-2017-09-806489.
51. Bödör, C. *et al.* (2012) 'Germ-line GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival', *Haematologica*. Ferrata Storti Foundation, 97(6), pp. 890–4. doi: 10.3324/haematol.2011.054361.
52. de Boer, J. *et al.* (2003) 'Transgenic mice with hematopoietic and lymphoid specific expression of Cre', *European Journal of Immunology*. Eur J Immunol, pp. 314–325. doi: 10.1002/immu.200310005.
53. Bogaert, D. J. *et al.* (2020) 'GATA2 deficiency and haematopoietic stem cell transplantation: challenges for the clinical practitioner', *British Journal of Haematology*. Blackwell Publishing Ltd, 188(5), pp. 768–773. doi: 10.1111/bjh.16247.
54. Borthakur, G. and Estey, E. E. (2007) 'Therapy-related acute myelogenous leukemia and myelodysplastic syndrome', *Current Oncology Reports*. Curr Oncol Rep, pp. 373–377. doi: 10.1007/s11912-007-0050-z.
55. Bothmer, A. *et al.* (2010) '53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination', *Journal of Experimental Medicine*. J Exp Med, 207(4), pp. 855–865. doi: 10.1084/jem.20100244.
56. Boyer, L. a L. a. *et al.* (2010) 'Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells', *Young*, 122(6), pp. 947–956. doi: 10.1016/j.cell.2005.08.020.Core.
57. De Braekeleer, E. *et al.* (2011) 'RUNX1 translocations and fusion genes in malignant hemopathies', *Future Oncology*, 7(1), pp. 77–91. doi: 10.2217/fon.10.158.
58. Brambilla, L. *et al.* (2015) 'Hitting the right spot: Mechanism of action of OPB-31121, a novel and potent inhibitor of the Signal Transducer and Activator of Transcription 3 (STAT3)', *Molecular Oncology*. Elsevier, 9(6), pp. 1194–1206. doi: 10.1016/j.molonc.2015.02.012.
59. Brambrink, T. *et al.* (2008) 'Sequential Expression of Pluripotency Markers during Direct Reprogramming of Mouse Somatic Cells', *Cell Stem Cell*, 2(2), pp. 151–159. doi: 10.1016/j.stem.2008.01.004.
60. Brehm, M. A. *et al.* (2012) 'Engraftment of human HSCs in nonirradiated newborn NOD-scid IL2rγ null mice is enhanced by transgenic expression of membrane-bound human SCF.', *Blood*. American Society of Hematology, 119(12), pp. 2778–88. doi: 10.1182/blood-2011-05-353243.
61. Broxmeyer, H. E. *et al.* (2020) 'Fate of Hematopoiesis During Aging. What Do We Really Know, and What are its Implications?', *Stem Cell Reviews and Reports*. Springer, pp. 1020–1048. doi: 10.1007/s12015-020-10065-y.
62. Bruijn, M. de and Dzierzak, E. (2017) 'Runx transcription factors in the development and function of the definitive hematopoietic system', *Blood*. American Society of Hematology, 129(15), pp. 2061–2069. doi: 10.1182/BLOOD-2016-12-689109.
63. De Bruin, A. M., Voermans, C. and Nolte, M. A. (2014) 'Impact of interferon-γ on hematopoiesis', *Blood*. doi: 10.1182/blood-2014-04-568451.
64. Buenrostro, J. D. *et al.* (2018) 'Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation', *Cell*. Cell Press, 173(6), pp. 1535-1548.e16. doi: 10.1016/j.cell.2018.03.074.
65. Burridge, P. W. *et al.* (2011) 'A Universal System for Highly Efficient Cardiac Differentiation of Human Induced Pluripotent Stem Cells That Eliminates Interline Variability', *PLoS ONE*. Edited by M. Pera. Public Library of Science, 6(4), p. e18293. doi: 10.1371/journal.pone.0018293.

66. Buza-Vidas, N. *et al.* (2011) 'GATA3 is redundant for maintenance and self-renewal of hematopoietic stem cells', *Blood*. Blood, 118(5), pp. 1291–1293. doi: 10.1182/blood-2011-02-338046.
67. Cabezas-Wallscheid, N. *et al.* (2013) 'Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model', *EMBO Molecular Medicine*. Wiley-Blackwell, 5(12), pp. 1804–1820. doi: 10.1002/emmm.201302661.
68. Cabezas-Wallscheid, N. *et al.* (2014) 'Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis', *Cell Stem Cell*. Cell Press, 15(4), pp. 507–522. doi: 10.1016/j.stem.2014.07.005.
69. Cai, S. F. and Levine, R. L. (2019) 'Genetic and epigenetic determinants of AML pathogenesis', *Seminars in Hematology*. Elsevier Inc., 56(2), pp. 84–89. doi: 10.1053/j.seminhematol.2018.08.001.
70. Cai, Z. *et al.* (2018) 'Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis', *Cell Stem Cell*. Cell Press, 23(6), pp. 833–849. doi: 10.1016/j.stem.2018.10.013.
71. Caisander, G. *et al.* (2006) 'Chromosomal integrity maintained in five human embryonic stem cell lines after prolonged in vitro culture', *Chromosome Research*. Springer, 14(2), pp. 131–137. doi: 10.1007/s10577-006-1019-8.
72. Calkhoven, C. F., Müller, C. and Leutz, A. (2000) 'Translational control of C/EBPalpha and C/EBPbeta isoform expression.', *Genes & development*. Cold Spring Harbor Laboratory Press, 14(15), pp. 1920–32. doi: 10.1101/gad.14.15.1920.
73. Calvi, L. M. and Link, D. C. (2014) 'Cellular complexity of the bone marrow hematopoietic stem cell niche', *Calcif Tissue Int.*, 94(1), pp. 112–124. doi: 10.1007/s00223-013-9805-8.Cellular.
74. Carrelha, J. *et al.* (2018) 'Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells', *Nature*. Nature Publishing Group, 554(7690), pp. 106–111. doi: 10.1038/nature25455.
75. Castaño, J. *et al.* (2019) 'GATA2 Promotes Hematopoietic Development and Represses Cardiac Differentiation of Human Mesoderm', *Stem Cell Reports*, 13. doi: 10.1016/j.stemcr.2019.07.009.
76. Cavalcante de Andrade Silva, M. *et al.* (2020) 'Breaking the spatial constraint between neighboring zinc fingers: a new germline mutation in GATA2 deficiency syndrome', *Leukemia*. doi: 10.1038/s41375-020-0820-2.
77. Cerdan, C., Rouleau, A. and Bhatia, M. (2004) 'VEGF-A165 augments erythropoietic development from human embryonic stem cells', *Blood*. Blood, 103(7), pp. 2504–2512. doi: 10.1182/blood-2003-07-2563.
78. Chadwick, K. *et al.* (2003) 'Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells', *Blood*, 102(3), pp. 906–915. doi: 10.1182/blood-2003-03-0832.
79. Chang, A. N. *et al.* (2002) 'GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 99(14), pp. 9237–9242. doi: 10.1073/pnas.142302099.
80. Chao, M. P. *et al.* (2017) 'Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease', *Cell Stem Cell*, 20(3), pp. 329-344.e7. doi: 10.1016/j.stem.2016.11.018.
81. Chao, M. P. and Majeti, R. (2019) 'Induced pluripotent stem cell modeling of malignant hematopoiesis', *Experimental Hematology*. Elsevier Inc., 71, pp. 68–76. doi: 10.1016/j.exphem.2019.01.002.
82. Chapman, J. R., Taylor, M. R. G. and Boulton, S. J. (2012) 'Playing the End Game: DNA Double-Strand Break Repair Pathway Choice', *Molecular Cell*. Elsevier, pp. 497–510. doi: 10.1016/j.molcel.2012.07.029.
83. Charles, M. A. *et al.* (2006) 'Pituitary-specific Gata2 knockout: Effects on gonadotrope and thyrotrope function', *Molecular Endocrinology*, 20(6), pp. 1366–1377. doi: 10.1210/me.2005-0378.
84. Chatzinikolaou, G., Karakasilioti, I. and Garinis, G. A. (2014) 'DNA damage and innate immunity: Links and trade-offs', *Trends in Immunology*. Elsevier Ltd, pp. 429–435. doi: 10.1016/j.it.2014.06.003.
85. Chen, J. *et al.* (2008) 'Enrichment of hematopoietic stem cells with SLAM and LSK markers for the detection of hematopoietic stem cell function in normal and Trp53 null mice', *Experimental Hematology*. NIH Public Access, 36(10), pp. 1236–1243. doi: 10.1016/j.exphem.2008.04.012.
86. Chen, M. J. *et al.* (2009) 'Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter', *Nature*. Nature Publishing Group, 457(7231), pp. 887–891. doi: 10.1038/nature07619.
87. Chen, X. *et al.* (no date) 'Bone marrow myeloid cells regulate myeloid-biased hematopoietic stem cells via a histamine-dependent feedback loop', *Elsevier*. Available at: https://www.sciencedirect.com/science/article/pii/S1934590917304551 (Accessed: 24 June 2020).
88. Chia, N. Y. *et al.* (2010) 'A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity', *Nature*. Nature Publishing Group, 468(7321), pp. 316–320. doi: 10.1038/nature09531.

89. Chicha, L. *et al.* (2011) 'Human Pluripotent Stem Cells Differentiated in Fully Defined Medium Generate Hematopoietic CD34 Progenitors with Distinct Characteristics Human Pluripotent Stem Cells Differentiated in Fully Defined Medium Generate Hematopoietic CD34 + and CD34'. doi: 10.1371/journal.pone.0014733.
90. Choi, K.-D. *et al.* (2009) 'Hematopoietic and endothelial differentiation of human induced pluripotent stem cells.', *Stem cells (Dayton, Ohio)*. NIH Public Access, 27(3), pp. 559–67. doi: 10.1634/stemcells.2008-0922.
91. Choi, K.-D., Vodyanik, M. A. and Slukvin, I. I. (2011) 'The Hematopoietic Differentiation and Production of Mature Myeloid Cells from Human Pluripotent Stem Cells', *Nature protocols*. Nature Publishing Group, 6(3), pp. 296–313. doi: 10.1038/nprot.2010.184.
92. Choi, K. D. *et al.* (2012) 'Identification of the Hemogenic Endothelial Progenitor and Its Direct Precursor in Human Pluripotent Stem Cell Differentiation Cultures', *Cell Reports*, 2(3), pp. 553–567. doi: 10.1016/j.celrep.2012.08.002.
93. Chong, C. E. *et al.* (2018) 'Differential effects on gene transcription and hematopoietic differentiation correlate with GATA2 mutant disease phenotypes', *Leukemia*. Nature Publishing Group, 32(1), pp. 194–202. doi: 10.1038/leu.2017.196.
94. Chou, W.-C. *et al.* (2010) 'Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations', *Blood*. American Society of Hematology, 116(20), pp. 4086–4094. doi: 10.1182/BLOOD-2010-05-283291.
95. Chow, A. *et al.* (2011) 'Bone marrow CD169+ macrophages promote the retention of hematopoietic stem and progenitor cells in the mesenchymal stem cell niche', *Journal of Experimental Medicine*. Rockefeller University Press, 208(2), pp. 761–771. doi: 10.1084/jem.20101688.
96. Churpek, J. E. *et al.* (2015) 'Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia', *Blood*. American Society of Hematology, 126(22), pp. 2484–2490. doi: 10.1182/blood-2015-04-641100.
97. Churpek, J. E. (2017) 'Familial myelodysplastic syndrome/acute myeloid leukemia', *Best Practice and Research: Clinical Haematology*. Baillière Tindall, pp. 287–289. doi: 10.1016/j.beha.2017.10.002.
98. Churpek, J. E. and Bresnick, E. H. (2019a) 'Transcription factor mutations as a cause of familial myeloid neoplasms', *Journal of Clinical Investigation*. American Society for Clinical Investigation, pp. 476–488. doi: 10.1172/JCI120854.
99. Churpek, J. E. and Bresnick, E. H. (2019b) 'Transcription factor mutations as a cause of familial myeloid neoplasms', *Journal of Clinical Investigation*. American Society for Clinical Investigation, pp. 476–488. doi: 10.1172/JCI120854.
100. Chute, J. P. *et al.* (2002) 'Rescue of Hematopoietic Stem Cells Following High-Dose Radiation Injury Using Ex Vivo Culture on Endothelial Monolayers', *Military Medicine*. doi: 10.1093/milmed/167.suppl_1.74.
101. Chute, J. P. *et al.* (2004) 'Ex vivo culture rescues hematopoietic stem cells with long-term repopulating capacity following harvest from lethally irradiated mice', *Experimental Hematology*. Exp Hematol, 32(3), pp. 308–317. doi: 10.1016/j.exphem.2003.12.002.
102. Ciesielska, A., Matyjek, M. and Kwiatkowska, K. (2020) 'TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling', *Cellular and Molecular Life Sciences*. Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s00018-020-03656-y.
103. Ciullini Mannurita, S. *et al.* (2016) 'Timely follow-up of a GATA2 deficiency patient allows successful treatment', *Journal of Allergy and Clinical Immunology*. Mosby Inc., 138(5), pp. 1480-1483.e4. doi: 10.1016/j.jaci.2016.06.004.
104. Clapes, T., Lefkopoulos, S. and Trompouki, E. (2016a) 'Stress and non-stress roles of inflammatory signals during HSC emergence and maintenance', *Frontiers in Immunology*. doi: 10.3389/fimmu.2016.00487.
105. Clapes, T., Lefkopoulos, S. and Trompouki, E. (2016b) 'Stress and non-stress roles of inflammatory signals during HSC emergence and maintenance', *Frontiers in Immunology*. Frontiers Media S.A., p. 1. doi: 10.3389/fimmu.2016.00487.
106. Collin, M., Dickinson, R. and Bigley, V. (2015a) 'Haematopoietic and immune defects associated with GATA2 mutation', *British Journal of Haematology*, pp. 173–187. doi: 10.1111/bjh.13317.
107. Collin, M., Dickinson, R. and Bigley, V. (2015b) 'Haematopoietic and immune defects associated with *GATA2* mutation', *British Journal of Haematology*, 169(2), pp. 173–187. doi: 10.1111/bjh.13317.
108. Connelly, J. P. *et al.* (2014) 'Targeted correction of RUNX1 mutation in FPD patient-specific induced pluripotent stem cells rescues megakaryopoietic defects', *Blood*. American Society of Hematology, 124(12), pp. 1926–1930. doi: 10.1182/blood-2014-01-550525.

109. Cook, E. K., Luo, M. and Rauh, M. J. (2020) 'Clonal hematopoiesis and inflammation: Partners in leukemogenesis and comorbidity', *Experimental Hematology*. Elsevier Inc., pp. 85–94. doi: 10.1016/j.exphem.2020.01.011.
110. Corat, M. A. F. *et al.* (2017) 'Acquired somatic mutations in PNH reveal long-term maintenance of adaptive NK cells independent of HSPCs', *Blood*. American Society of Hematology, 129(14), pp. 1940–1946. doi: 10.1182/blood-2016-08-734285.
111. Cordeiro Gomes, A. *et al.* (2016) 'Hematopoietic Stem Cell Niches Produce Lineage-Instructive Signals to Control Multipotent Progenitor Differentiation', *Immunity*. Cell Press, 45(6), pp. 1219–1231. doi: 10.1016/j.immuni.2016.11.004.
112. Cormican, S. and Griffin, M. D. (2020) 'Human Monocyte Subset Distinctions and Function: Insights From Gene Expression Analysis', *Frontiers in Immunology*, p. 1070. doi: 10.3389/fimmu.2020.01070.
113. Cortes-Lavaud, X. *et al.* (2015) 'GATA2 Germline Mutations Impair GATA2 Transcription, Causing Haploinsufficiency: Functional Analysis of the p.Arg396Gln Mutation', *The Journal of Immunology*, 194(5), pp. 2190–2198. doi: 10.4049/jimmunol.1401868.
114. Coussens, L. M. and Werb, Z. (2002) 'Inflammation and cancer', *Nature*. Nature, pp. 860–867. doi: 10.1038/nature01322.
115. Cros, J. *et al.* (2010) 'Human CD14dim Monocytes Patrol and Sense Nucleic Acids and Viruses via TLR7 and TLR8 Receptors', *Immunity*. Immunity, 33(3), pp. 375–386. doi: 10.1016/j.immuni.2010.08.012.
116. Cuellar-Rodriguez, J. *et al.* (2011) 'Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency', *Blood*. doi: 10.1182/blood-2011-06-365049.
117. Cull, A. H. *et al.* (2017a) 'Tet2 restrains inflammatory gene expression in macrophages', *Experimental Hematology*. Elsevier Inc., 55, pp. 56-70.e13. doi: 10.1016/j.exphem.2017.08.001.
118. Cull, A. H. *et al.* (2017b) 'Tet2 restrains inflammatory gene expression in macrophages', *Experimental Hematology*. Elsevier Inc., 55, pp. 56-70.e13. doi: 10.1016/j.exphem.2017.08.001.
119. Cutler, C. S. *et al.* (2004) 'A decision analysis of allogeneic bone marrow transplantation for the myelodysplastic syndromes: Delayed transplantation for low-risk myelodysplasia is associated with improved outcome', *Blood*. Elsevier, 104(2), pp. 579–585. doi: 10.1182/blood-2004-01-0338.
120. Czuchlewski, D. R. and Peterson, L. A. C. (2016) 'Myeloid Neoplasms with Germline Predisposition. A New Provisional Entity Within the World Health Organization Classification.', *Surgical Pathology Clinics*. W.B. Saunders, pp. 165–176. doi: 10.1016/j.path.2015.09.010.
121. David, L. and Polo, J. M. (2014) 'Phases of reprogramming', *Stem Cell Research*. Elsevier, pp. 754–761. doi: 10.1016/j.scr.2014.03.007.
122. Delmonte, O. M. *et al.* (2019) 'Inborn Errors of Immunity With Immune Dysregulation: From Bench to Bedside', *Frontiers in Pediatrics*. Frontiers Media S.A., p. 353. doi: 10.3389/fped.2019.00353.
123. Demchenko, A. P. (2013) 'Beyond annexin V: Fluorescence response of cellular membranes to apoptosis', *Cytotechnology*. Springer, pp. 157–172. doi: 10.1007/s10616-012-9481-y.
124. DeWitt, M. A., Corn, J. E. and Carroll, D. (2017) 'Genome editing via delivery of Cas9 ribonucleoprotein', *Methods*. Academic Press Inc., 121–122, pp. 9–15. doi: 10.1016/j.ymeth.2017.04.003.
125. Dickinson, R. E. *et al.* (2011) 'Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency', *Blood*, 118(10). doi: 10.1182/blood-2011-06-360313.
126. Dickinson, R. E. *et al.* (2014) 'The evolution of cellular deficiency in GATA2 mutation', *Blood*, 123(6), pp. 863–874. doi: 10.1182/blood-2013-07-517151.
127. DiMascio, L. *et al.* (2007) 'Identification of Adiponectin as a Novel Hemopoietic Stem Cell Growth Factor', *The Journal of Immunology*. The American Association of Immunologists, 178(6), pp. 3511–3520. doi: 10.4049/jimmunol.178.6.3511.
128. Ding, L. *et al.* (2012) 'Endothelial and perivascular cells maintain haematopoietic stem cells', *Nature*. Nature Publishing Group, 481(7382), pp. 457–462. doi: 10.1038/nature10783.
129. Ding, L. and Morrison, S. J. (2013) 'Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches', *Nature*. Nature, 495(7440), pp. 231–235. doi: 10.1038/nature11885.
130. Ditadi, A. *et al.* (2015) 'Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages', *Nature Cell Biology*. Nature Publishing Group, 17(5), pp. 580–591. doi: 10.1038/ncb3161.
131. Djuretic, I. M., Cruz-Guilloty, F. and Rao, A. (2009) 'Regulation of gene expression in peripheral T cells by Runx transcription factors', *Advances in Immunology*. Academic Press Inc., 104(C), pp. 1–23. doi: 10.1016/S0065-2776(08)04001-7.
132. Doan, P. L. *et al.* (2013) 'Tie21 bone marrow endothelial cells regulate hematopoietic stem cell regeneration following radiation injury', *Stem Cells*. Stem Cells, 31(2), pp. 327–337. doi:

10.1002/stem.1275.
133. Doench, J. G. *et al.* (2016) 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', *Nature Biotechnology*. Nature Publishing Group, 34(2), pp. 184–191. doi: 10.1038/nbt.3437.
134. Donadieu, J. *et al.* (2018) 'Natural history of GATA2 deficiency in a survey of 79 French and Belgian patients', *Haematologica*. Ferrata Storti Foundation, 103(8), pp. 1278–1287. doi: 10.3324/haematol.2017.181909.
135. Doulatov, S. *et al.* (2010) 'Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development', *Nature Immunology*. Nature Publishing Group, 11(7), pp. 585–593. doi: 10.1038/ni.1889.
136. Doulatov, S., Notta, F., Laurenti, E. and Dick, J. E. (2012) 'Cell Stem Cell Hematopoiesis: A Human Perspective'. doi: 10.1016/j.stem.2012.01.006.
137. Doulatov, S., Notta, F., Laurenti, E. and Dick, J. E. E. (2012) 'Hematopoiesis: A Human Perspective', *Cell Stem Cell*. Cell Press, 10(2), pp. 120–136. doi: 10.1016/J.STEM.2012.01.006.
138. Doulatov, S. and Papapetrou, E. P. (2021) 'Studying clonal evolution of myeloid malignancies using induced pluripotent stem cells', *Current opinion in hematology*. NLM (Medline), 28(1), pp. 50–56. doi: 10.1097/MOH.0000000000000620.
139. Draper, J. E. *et al.* (2017) 'Mouse RUNX1C regulates premegakaryocytic/erythroid output and maintains survival of megakaryocyte progenitors', *Blood*. American Society of Hematology, 130(3), pp. 271–284. doi: 10.1182/blood-2016-06-723635.
140. Dzierzak, E. and Speck, N. a (2008) 'Of lineage and legacy: the development of mammalian hematopoietic stem cells.', *Nature immunology*, 9(2), pp. 129–136. doi: 10.1038/ni1560.
141. Ebert, B. L. *et al.* (2008) 'Identification of RPS14 as a 5q- syndrome gene by RNA interference screen', *Nature*. Nature Publishing Group, 451(7176), pp. 335–339. doi: 10.1038/nature06494.
142. Van Eijkelenburg, N. K. A. *et al.* (2018) 'Clofarabine, high-dose cytarabine and liposomal daunorubicin in pediatric relapsed/refractory acute myeloid leukemia: A phase IB study', *Haematologica*. Ferrata Storti Foundation, 103(9), pp. 1484–1492. doi: 10.3324/haematol.2017.187153.
143. Esteban, M. A. *et al.* (2012) 'The mesenchymal-to-epithelial transition in somatic cell reprogramming', *Current Opinion in Genetics and Development*. Elsevier Current Trends, pp. 423–428. doi: 10.1016/j.gde.2012.09.004.
144. Ezoe, S. *et al.* (2002) 'GATA-2/estrogen receptor chimera regulates cytokine-dependent growth of hematopoietic cells through accumulation of p21WAF1 and p27Kip1 proteins', *Blood*. Blood, 100(10), pp. 3512–3520. doi: 10.1182/blood-2002-04-1177.
145. Fan, A. *et al.* (2020) 'Cytokine Rescue and Targeting of Inflammation-Sensitive RUNX1 Deficient Human CD34+ Hematopoietic Stem and Progenitor Cells'. ASH.
146. Fang, J. *et al.* (2017) 'Ubiquitination of hnRNPA1 by TRAF6 links chronic innate immune signaling with myelodysplasia', *Nature Immunology*. Nature Publishing Group, 18(2), pp. 236–245. doi: 10.1038/ni.3654.
147. Fasan, A. *et al.* (2013) 'GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis', *Leukemia*. Nature Publishing Group, 27(2), pp. 482–485. doi: 10.1038/leu.2012.174.
148. Fawaz, S. *et al.* (2021) 'Clonal haematopoiesis and cardiovascular diseases: A growing relationship.', *Archives of cardiovascular diseases*. Arch Cardiovasc Dis. doi: 10.1016/j.acvd.2021.01.002.
149. Fehse, B. *et al.* (2004) 'Pois(s)on – It's a Question of Dose…', *Gene Therapy*. Nature Publishing Group, 11(11), pp. 879–881. doi: 10.1038/sj.gt.3302270.
150. Féraud, O. *et al.* (2016) 'Donor dependent variations in hematopoietic differentiation among embryonic and induced pluripotent stem cell lines', *PLoS ONE*. Edited by D. Kaufman. Public Library of Science, 11(3), p. e0149291. doi: 10.1371/journal.pone.0149291.
151. Ferrando, A. A. *et al.* (2002) 'Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia', *Cancer Cell*, 1(1), pp. 75–87. doi: 10.1016/S1535-6108(02)00018-1.
152. Ferrando, A. A. *et al.* (2004) 'Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia', *Blood*, 103(5), pp. 1909–1911. doi: 10.1182/blood-2003-07-2577.
153. Ferraro, F. *et al.* (2011) 'Diabetes impairs hematopoietic stem cell mobilization by altering niche function', *Science Translational Medicine*. American Association for the Advancement of Science, 3(104). doi: 10.1126/scitranslmed.3002191.
154. Feurstein, S. and Godley, L. A. (2017) 'Germline ETV6 mutations and predisposition to hematological malignancies', *International Journal of Hematology*. Springer Tokyo, pp. 189–195. doi: 10.1007/s12185-017-2259-4.

155. Fiedler, K. and Brunner, C. (2012) 'The role of transcription factors in the guidance of granulopoiesis.', *American journal of blood research*. e-Century Publishing Corporation, 2(1), pp. 57–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22432088 (Accessed: 26 March 2021).
156. Figueroa, M. E. *et al.* (2009) 'MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation', *Blood*. Blood, 114(16), pp. 3448–3458. doi: 10.1182/blood-2009-01-200519.
157. Frame, J. M., McGrath, K. E. and Palis, J. (2013) 'Erythro-myeloid progenitors: "Definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells', *Blood Cells, Molecules, and Diseases*, pp. 220–225. doi: 10.1016/j.bcmd.2013.09.006.
158. Frelin, C. *et al.* (2013) 'GATA-3 regulates the self-renewal of long-term hematopoietic stem cells', *Nature Immunology*. PMC Canada manuscript submission, 14(10), pp. 1037–1044. doi: 10.1038/ni.2692.
159. Friedman, A. D. (2007) 'Transcriptional control of granulocyte and monocyte development', *Oncogene*. Oncogene, pp. 6816–6828. doi: 10.1038/sj.onc.1210764.
160. Fujino, T. and Kitamura, T. (2020) 'ASXL1 mutation in clonal hematopoiesis', *Experimental Hematology*, 83, pp. 74–84. doi: 10.1016/j.exphem.2020.01.002.
161. Fujiwara, T. *et al.* (2009) 'Discovering Hematopoietic Mechanisms through Genome-wide Analysis of GATA Factor Chromatin Occupancy', *Molecular Cell*. Cell Press, 36(4), pp. 667–681. doi: 10.1016/j.molcel.2009.11.001.
162. Fuster, J. J. *et al.* (2017) 'Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice', *Science*. American Association for the Advancement of Science, 355(6327), pp. 842–847. doi: 10.1126/science.aag1381.
163. Gafni, O. *et al.* (2013) 'Derivation of novel human ground state naive pluripotent stem cells', *Nature*. Nature Publishing Group, 504(7479), pp. 282–286. doi: 10.1038/nature12745.
164. Galán-Díez, M. and Kousteni, S. (2018) 'A bone marrow niche-derived molecular switch between osteogenesis and hematopoiesis', *Genes and Development*. Cold Spring Harbor Laboratory Press, 32(5–6), pp. 324–326. doi: 10.1101/gad.314013.118.
165. Galera, P. *et al.* (2018) 'Donor-derived MDS/AML in families with germline GATA2 mutation', *Blood*, pp. 1994–1998. doi: 10.1182/blood-2018-07-861070.
166. Galera, P., Dulau-Florea, A. and Calvo, K. R. (2019) 'Inherited thrombocytopenia and platelet disorders with germline predisposition to myeloid neoplasia', *International Journal of Laboratory Hematology*. Blackwell Publishing Ltd, pp. 131–141. doi: 10.1111/ijlh.12999.
167. Ganapathi, K. A. *et al.* (2015) 'GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia', *Blood*. American Society of Hematology, 125(1), pp. 56–70. doi: 10.1182/blood-2014-06-580340.
168. Gao, X. *et al.* (2013) 'Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo', *Journal of Experimental Medicine*. The Rockefeller University Press, 210(13), pp. 2833–2842. doi: 10.1084/jem.20130733.
169. Garcia-Manero, G. *et al.* (2019) 'Pracinostat plus azacitidine in older patients with newly diagnosed acute myeloid leukemia: Results of a phase 2 study', *Blood Advances*. American Society of Hematology, 3(4), pp. 508–518. doi: 10.1182/bloodadvances.2018027409.
170. Garćia-Ojeda, M. E. *et al.* (2013) 'GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice', *Blood*. American Society of Hematology, 121(10), pp. 1749–1759. doi: 10.1182/blood-2012-06-440065.
171. Garneau, J. E. *et al.* (2010) 'The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA', *Nature*. Nature Publishing Group, 468(7320), pp. 67–71. doi: 10.1038/nature09523.
172. Gaudelli, N. M. *et al.* (2017) 'Programmable base editing of A • T to G • C in genomic DNA without DNA cleavage', *Nature Publishing Group*, 551(7681), pp. 464–471. doi: 10.1038/nature24644.
173. Gehrke, J. M. *et al.* (2018) 'An apobec3a-cas9 base editor with minimized bystander and off-target activities', *Nature Biotechnology*. Nature Publishing Group, 36(10), p. 977. doi: 10.1038/nbt.4199.
174. Gekas, C. *et al.* (2005) 'The placenta is a niche for hematopoietic stem cells', *Developmental Cell*. Elsevier, 8(3), pp. 365–375. doi: 10.1016/j.devcel.2004.12.016.
175. Gelsi-Boyer, V. *et al.* (2009) 'Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia', *British Journal of Haematology*. John Wiley & Sons, Ltd, 145(6), pp. 788–800. doi: 10.1111/j.1365-2141.2009.07697.x.
176. Genovese, G. *et al.* (2014) 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence', *New England Journal of Medicine*. Massachussetts Medical Society, 371(26), pp. 2477–2487. doi: 10.1056/NEJMoa1409405.
177. Gerby, B. *et al.* (2014) 'SCL, LMO1 and Notch1 Reprogram Thymocytes into Self-Renewing Cells',

PLoS Genetics. Edited by H. L. Grimes and G. P. Copenhaver. Public Library of Science, 10(12), p. e1004768. doi: 10.1371/journal.pgen.1004768.
178. Godley, L. A. (2014) 'Inherited Predisposition to Acute Myeloid Leukemia', *Seminars in Hematology*, 51(4), pp. 306–321. doi: 10.1053/j.seminhematol.2014.08.001.
179. Godley, L. A. and Larson, R. A. (2008) 'Therapy-Related Myeloid Leukemia', *Seminars in Oncology*. Semin Oncol, 35(4), pp. 418–429. doi: 10.1053/j.seminoncol.2008.04.012.
180. Goldfarb, A. N. (2009) 'Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb', *Journal of Cellular Biochemistry*. NIH Public Access, pp. 377–382. doi: 10.1002/jcb.22142.
181. Golub, D. *et al.* (2019) 'Mutant isocitrate dehydrogenase inhibitors as targeted cancer therapeutics', *Frontiers in Oncology*. Frontiers Media S.A., 9(MAY). doi: 10.3389/fonc.2019.00417.
182. Golub, T. R. *et al.* (1994) 'Fusion of PDGF receptor β to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation', *Cell*. Cell, 77(2), pp. 307–316. doi: 10.1016/0092-8674(94)90322-0.
183. Govindarajah, V. and Reynaud, D. (2018) 'Tuning of the Hematopoietic Stem Cell Compartment in its Inflammatory Environment', *Current Stem Cell Reports*. Springer International Publishing, pp. 189–200. doi: 10.1007/s40778-018-0131-y.
184. Grandela, C. and Wolvetang, E. (2007) 'HESC adaptation, selection and stability', *Stem Cell Reviews*. Springer, 3(3), pp. 183–191. doi: 10.1007/s12015-007-0008-4.
185. Green, C. L. *et al.* (2013) 'GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations', *British Journal of Haematology*, 161(5), pp. 701–705. doi: 10.1111/bjh.12317.
186. Greenbaum, A. *et al.* (2013) 'CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance', *Nature*. Nature, 495(7440), pp. 227–230. doi: 10.1038/nature11926.
187. Greenberg, P. L. *et al.* (2012) 'Revised international prognostic scoring system for myelodysplastic syndromes', *Blood*. The American Society of Hematology, 120(12), pp. 2454–2465. doi: 10.1182/blood-2012-03-420489.
188. Greif, P. A. *et al.* (2012) 'GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia', *Blood*. American Society of Hematology, 120(2), pp. 395–403. doi: 10.1182/blood-2012-01-403220.
189. Grisolano, J. L. *et al.* (2003) 'An activated receptor tyrosine kinase, TEL/PDGFβR, cooperates with AML1/ETO to induce acute myeloid leukemia in mice', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 100(16), pp. 9506–9511. doi: 10.1073/pnas.1531730100.
190. Grote, D. *et al.* (2008) 'Gata3 acts downstream of β-catenin signaling to prevent ectopic metanephric kidney induction', *PLoS Genetics*. Public Library of Science, 4(12). doi: 10.1371/journal.pgen.1000316.
191. Grünewald, J. *et al.* (2019) 'Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors', *Nature*. Nature Publishing Group, 569(7756), pp. 433–437. doi: 10.1038/s41586-019-1161-z.
192. Gu, K. L. *et al.* (2016) 'Pluripotency-associated miR-290/302 family of microRNAs promote the dismantling of naive pluripotency', *Cell Research*. Nature Publishing Group, 26(3), pp. 350–366. doi: 10.1038/cr.2016.2.
193. Guha, M. and Mackman, N. (2001) 'LPS induction of gene expression in human monocytes', *Cellular Signalling*. Pergamon, pp. 85–94. doi: 10.1016/S0898-6568(00)00149-2.
194. Guo, G. *et al.* (2009) 'Klf4 reverts developmentally programmed restriction of ground state pluripotency', *Development*. The Company of Biologists Ltd, 136(7), pp. 1063–1069. doi: 10.1242/dev.030957.
195. Guo, P. *et al.* (2017) 'Endothelial jagged-2 sustains hematopoietic stem and progenitor reconstitution after myelosuppression', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 127(12), pp. 4242–4256. doi: 10.1172/JCI92309.
196. Gupta, R. *et al.* (2018) 'DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity', *Cell*. Cell Press, 173(4), pp. 972-988.e23. doi: 10.1016/j.cell.2018.03.050.
197. Guryanova, O. A. *et al.* (2016) 'DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling', *Nature Medicine*. Nature Publishing Group, 22(12), pp. 1488–1495. doi: 10.1038/nm.4210.
198. Haase, D. *et al.* (2007) 'New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: Evidence from a core dataset of 2124 patients', *Blood*. American Society of Hematology, 110(13), pp. 4385–4395. doi: 10.1182/blood-2007-03-082404.
199. Haeussler, M. *et al.* (2016) 'Evaluation of off-target and on-target scoring algorithms and integration into

the guide RNA selection tool CRISPOR', *Genome Biology*. BioMed Central Ltd., 17(1), p. 148. doi: 10.1186/s13059-016-1012-2.
200. Hahn, C. N. *et al.* (2011) 'Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia', *Nature genetics*. Nature Publishing Group, 43(10), pp. 1012–1017. doi: 10.1038/ng.913.
201. Hanna, J. *et al.* (2010) 'Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(20), pp. 9222–9227. doi: 10.1073/pnas.1004584107.
202. Haridhasapavalan, K. K. *et al.* (2020) 'An Insight into Reprogramming Barriers to iPSC Generation', *Stem Cell Reviews and Reports*. Springer, pp. 56–81. doi: 10.1007/s12015-019-09931-1.
203. Harigae, H. (2006) 'GATA Transcription Factors and Hematological Diseases', *Tohoku J. Exp. Med*, 210(1), pp. 1–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16960339 (Accessed: 29 June 2017).
204. Hasemann, M. S. *et al.* (2014) 'C/EBPα Is Required for Long-Term Self-Renewal and Lineage Priming of Hematopoietic Stem Cells and for the Maintenance of Epigenetic Configurations in Multipotent Progenitors', *PLoS Genetics*. Edited by H. L. Grimes, 10(1), p. e1004079. doi: 10.1371/journal.pgen.1004079.
205. Hershberger, C. E. *et al.* (2016) 'The Role of LUC7L2 in Splicing and MDS', *Blood*. American Society of Hematology, 128(22), pp. 5504–5504. doi: 10.1182/blood.v128.22.5504.5504.
206. Herzenberg, Leonore A. *et al.* (2006) 'Interpreting flow cytometry data: A guide for the perplexed', *Nature Immunology*. Nature Publishing Group, pp. 681–685. doi: 10.1038/ni0706-681.
207. Hille, F. *et al.* (2018) 'The Biology of CRISPR-Cas: Backward and Forward', *Cell*. Cell Press, pp. 1239–1259. doi: 10.1016/j.cell.2017.11.032.
208. Hirai, H. *et al.* (2003) 'Hemogenic and nonhemogenic endothelium can be distinguished by the activity of fetal liver kinase (Flk)-1 promoter/enhancer during mouse embryogenesis', *Blood*. American Society of Hematology, 101(3), pp. 886–893. doi: 10.1182/blood-2002-02-0655.
209. Hock, H. *et al.* (2004) 'Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival', *Genes and Development*. Genes Dev, 18(19), pp. 2336–2341. doi: 10.1101/gad.1239604.
210. Hollenhorst, P. C., McIntosh, L. P. and Graves, B. J. (2011) 'Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors', *Annual Review of Biochemistry*. Annual Reviews, 80(1), pp. 437–471. doi: 10.1146/annurev.biochem.79.081507.103945.
211. Hormaechea-Agulla, D. *et al.* (2021) 'Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling', *Cell Stem Cell*. Cell Press, 28, pp. 1–15. doi: 10.1016/J.STEM.2021.03.002.
212. Hoshino, T. *et al.* (2008) 'Reduced BMP4 abundance in Gata2 hypomorphic mutant mice result in uropathies resembling human CAKUT', *Genes to Cells*. Genes Cells, 13(2), pp. 159–170. doi: 10.1111/j.1365-2443.2007.01158.x.
213. Hosono, N. *et al.* (2014) 'Recurrent genetic defects on chromosome 7q in myeloid neoplasms', *Leukemia*. Nature Publishing Group, pp. 1348–1351. doi: 10.1038/leu.2014.25.
214. Hosono, N. (2019) 'Genetic abnormalities and pathophysiology of MDS', *International Journal of Clinical Oncology*. Springer Tokyo, pp. 885–892. doi: 10.1007/s10147-019-01462-6.
215. Hosseini, M. M. *et al.* (2018) 'Inhibition of interleukin-1 receptor-associated kinase-1 is a therapeutic strategy for acute myeloid leukemia subtypes', *Leukemia*. Nature Publishing Group, 32(11), pp. 2374–2387. doi: 10.1038/s41375-018-0112-2.
216. Hotta, A. and Ellis, J. (2008) 'Retroviral vector silencing during iPS cell induction: An epigenetic beacon that signals distinct pluripotent states', *Journal of Cellular Biochemistry*. J Cell Biochem, pp. 940–948. doi: 10.1002/jcb.21912.
217. Hoy, S. M. (2019) 'Glasdegib: First Global Approval', *Drugs*. Springer International Publishing, 79(2), pp. 207–213. doi: 10.1007/s40265-018-1047-7.
218. Hsu, A. P. *et al.* (2011) 'Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome', *Blood*, 118(10), pp. 2653–2655. doi: 10.1182/blood-2011-05-356352.
219. Hsu, A. P. *et al.* (2013a) 'GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome.', *Blood*. American Society of Hematology, 121(19), pp. 3830–7, S1-7. doi: 10.1182/blood-2012-08-452763.
220. Hsu, A. P. *et al.* (2013b) 'GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome', *Blood*. American Society of Hematology, 121(19), pp. 3830–3837. Available at: http://www.bloodjournal.org/content/bloodjournal/121/19/3830.full.pdf (Accessed: 18 July 2017).

221. Hsu, H. T. *et al.* (2015) 'Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4', *Science*. American Association for the Advancement of Science, 348(6241), pp. 1372–1376. doi: 10.1126/science.aab1223.
222. Hsu, J. I. *et al.* (2018) 'PPM1D Mutations Drive Clonal Hematopoiesis in Response to Cytotoxic Chemotherapy', *Cell Stem Cell*. Cell Press, 23(5), pp. 700-713.e6. doi: 10.1016/j.stem.2018.10.004.
223. Hyman, B. and Tanzi, R. E. (2019) 'Effects of Species-Specific Genetics on Alzheimer's Mouse Models', *Neuron*, pp. 351–352. doi: 10.1016/j.neuron.2019.01.021.
224. Inaba, T., Honda, H. and Matsui, H. (2018) 'The enigma of monosomy 7', *Blood*, 131(26), p. blood-2017-12-822262. doi: 10.1182/blood-2017-12-822262.
225. Inoue, D. *et al.* (2013) 'Myelodysplastic syndromes are induced by histone methylation-altering ASXL1 mutations', *Journal of Clinical Investigation*, 123(11), pp. 4627–4640. doi: 10.1172/JCI70739.
226. Iqbal, A. J., Fisher, E. A. and Greaves, D. R. (2016) 'Inflammation—a Critical Appreciation of the Role of Myeloid Cells', in *Myeloid Cells in Health and Disease*. American Society of Microbiology, pp. 325–342. doi: 10.1128/microbiolspec.mchd-0027-2016.
227. Itkin, T. *et al.* (2016) 'Distinct bone marrow blood vessels differentially regulate haematopoiesis', *Nature*. Nature Publishing Group, 532(7599), pp. 323–328. doi: 10.1038/nature17624.
228. Ivanovs, A. *et al.* (2014) 'Identification of the niche and phenotype of the first human hematopoietic stem cells', *Stem Cell Reports*. Cell Press, 2(4), pp. 449–456. doi: 10.1016/j.stemcr.2014.02.004.
229. Iwasaki, H. *et al.* (2006) 'The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages', *Genes and Development*. Genes Dev, 20(21), pp. 3010–3021. doi: 10.1101/gad.1493506.
230. Jaiswal, S. *et al.* (2014a) 'Age-related clonal hematopoiesis associated with adverse outcomes', *New England Journal of Medicine*. Massachussetts Medical Society, 371(26), pp. 2488–2498. doi: 10.1056/NEJMoa1408617.
231. Jaiswal, S. *et al.* (2014b) 'Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes', *New England Journal of Medicine*. Massachusetts Medical Society, 371(26), pp. 2488–2498. doi: 10.1056/nejmoa1408617.
232. Jaiswal, S. *et al.* (2017) 'Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease', *New England Journal of Medicine*. New England Journal of Medicine (NEJM/MMS), 377(2), pp. 111–121. doi: 10.1056/nejmoa1701719.
233. Jalagadugula, G. *et al.* (2010) 'Regulation of platelet myosin light chain (MYL9) by RUNX1: Implications for thrombocytopenia and platelet dysfunction in RUNX1 haplodeficiency', *Blood*. American Society of Hematology, 116(26), pp. 6037–6045. doi: 10.1182/blood-2010-06-289850.
234. Jalbert, E. and Pietras, E. M. (2018) 'Analysis of murine hematopoietic stem cell proliferation during inflammation', in *Methods in Molecular Biology*. Humana Press Inc., pp. 183–200. doi: 10.1007/978-1-4939-7371-2_14.
235. Jasin, M. and Haber, J. E. (2016) 'The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair', *DNA Repair*. Elsevier B.V., pp. 6–16. doi: 10.1016/j.dnarep.2016.05.001.
236. Jasin, M. and Rothstein, R. (2013) 'Repair of strand breaks by homologous recombination', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 5(11). doi: 10.1101/cshperspect.a012740.
237. Jerez, A. *et al.* (2012) 'Loss of heterozygosity in 7q myeloid disorders: Clinical associations and genomic pathogenesis', *Blood*. The American Society of Hematology, 119(25), pp. 6109–6117. doi: 10.1182/blood-2011-12-397620.
238. Jiang, F. *et al.* (2016) 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage', *Science*. American Association for the Advancement of Science, 351(6275), pp. 867–871. doi: 10.1126/science.aad8282.
239. Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*. American Association for the Advancement of Science, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.
240. Jinek, M. *et al.* (2014) 'Structures of Cas9 endonucleases reveal RNA-mediated conformational activation', *Science*. American Association for the Advancement of Science, 343(6176), pp. 1215–1227. doi: 10.1126/science.1247997.
241. Johnson, Kirby D *et al.* (2012) 'Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity.', *The Journal of clinical investigation*. American Society for Clinical Investigation, 122(10), pp. 3692–704. doi: 10.1172/JCI61623.
242. Johnson, Kirby D. *et al.* (2012) 'Cis-element mutated in GATA2-dependent immunodeficiency governs

hematopoiesis and vascular integrity', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 122(10), pp. 3692–3704. doi: 10.1172/JCI61623.
243. Johnson, K. D. *et al.* (2015) 'Cis-regulatory mechanisms governing stem and progenitor cell transitions', *Science Advances*. American Association for the Advancement of Science, 1(8), p. e1500503. doi: 10.1126/sciadv.1500503.
244. Jouneau, A. *et al.* (2012) 'Naive and primed murine pluripotent stem cells have distinct miRNA expression profiles', *RNA*. Cold Spring Harbor Laboratory Press, 18(2), pp. 253–264. doi: 10.1261/rna.028878.111.
245. Jung, M. *et al.* (2018) 'GATA2 deficiency and human hematopoietic development modeled using induced pluripotent stem cells', *Blood Advances*. The American Society of Hematology, 2(23), pp. 3553–3565. doi: 10.1182/bloodadvances.2018017137.
246. Jung, M., Dunbar, C. E. and Winkler, T. (2015) 'Modeling Human Bone Marrow Failure Syndromes Using Pluripotent Stem Cells and Genome Engineering', *Molecular Therapy*. Cell Press, 23(12), pp. 1832–1842. doi: 10.1038/MT.2015.180.
247. Kaimakis, P. *et al.* (2016) 'Functional and molecular characterization of mouse Gata2-independent hematopoietic progenitors', *Blood*. doi: 10.1182/blood-2015-10-673749.
248. Kang, H. J. *et al.* (2018) 'GATA2 Is Dispensable for Specification of Hemogenic Endothelium but Promotes Endothelial-to-Hematopoietic Transition', *Stem Cell Reports*. Cell Press, 11(1), pp. 197–211. doi: 10.1016/j.stemcr.2018.05.002.
249. Karamitros, D. *et al.* (2018) 'Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells article', *Nature Immunology*. Nature Publishing Group, 19(1), pp. 85–97. doi: 10.1038/s41590-017-0001-2.
250. Karin, M. (2006) 'Nuclear factor-κB in cancer development and progression', *Nature*. Nature, pp. 431–436. doi: 10.1038/nature04870.
251. Katsumura, K. R. *et al.* (2017) 'Review Series TRANSCRIPTION FACTORS IN HEMATOPOIESIS AND HEMATOLOGIC DISEASE The GATA factor revolution in hematology', 129(15), pp. 2092–2103. doi: 10.1182/blood-2016-09.
252. Katsumura, K. R. *et al.* (2018) 'Human leukemia mutations corrupt but do not abrogate GATA-2 function.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(43), pp. E10109–E10118. doi: 10.1073/pnas.1813015115.
253. Kaufman, C. K. *et al.* (2003) 'GATA-3: An unexpected regulator of cell lineage determination in skin', *Genes and Development*. Cold Spring Harbor Laboratory Press, 17(17), pp. 2108–2122. doi: 10.1101/gad.1115203.
254. Kaur, G. *et al.* (2010) 'RUNX1/core binding factor A2 regulates platelet 12-lipoxygenase gene (ALOX12): Studies in human RUNX1 haplodeficiency', *Blood*. American Society of Hematology, 115(15), pp. 3128–3135. doi: 10.1182/blood-2009-04-214601.
255. Kaustio, M. *et al.* (2021) 'Loss of DIAPH1 causes SCBMS, combined immunodeficiency and mitochondrial dysfunction', *Journal of Allergy and Clinical Immunology*. J Allergy Clin Immunol. doi: 10.1016/j.jaci.2020.12.656.
256. Kauts, M. L., Vink, C. S. and Dzierzak, E. (2016) 'Hematopoietic (stem) cell development — how divergent are the roads taken?', *FEBS Letters*, pp. 3975–3986. doi: 10.1002/1873-3468.12372.
257. Kawai, T. and Akira, S. (2008) 'Toll-like Receptor and RIG-1-like Receptor Signaling', *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc., 1143(1), pp. 1–20. doi: 10.1196/annals.1443.020.
258. Kayser, S. *et al.* (2011) 'The impact of therapy-related acute myeloid leukemia (AML) on outcome in 2853 adult patients with newly diagnosed AML', *Blood*. American Society of Hematology, 117(7), pp. 2137–2145. doi: 10.1182/blood-2010-08-301713.
259. Kellaway, S. *et al.* (2020) 't(8;21) Acute Myeloid Leukemia as a Paradigm for the Understanding of Leukemogenesis at the Level of Gene Regulation and Chromatin Programming', *Cells*. NLM (Medline). doi: 10.3390/cells9122681.
260. Kennedy, J. A. and Ebert, B. L. (2017) 'Clinical implications of Genetic mutations in Myelodysplastic syndrome', *Journal of Clinical Oncology*. American Society of Clinical Oncology, pp. 968–974. doi: 10.1200/JCO.2016.71.0806.
261. Khaled, M. *et al.* (2021) 'CRISPR/Cas9 mediated knock-out of VPREB1 gene induces a cytotoxic effect in myeloma cells', *PLoS ONE*. Public Library of Science, 16(1 January). doi: 10.1371/journal.pone.0245349.
262. Kiel, M. J. *et al.* (2005) 'SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells', *Cell*. Elsevier, 121(7), pp. 1109–1121. doi: 10.1016/j.cell.2005.05.026.

263. Kim, N. *et al.* (2019) 'Monozygotic twins with shared de novo GATA2 mutation but dissimilar phenotypes due to differential promoter methylation', *Leukemia and Lymphoma*. Taylor and Francis Ltd, 60(4), pp. 1053–1061. doi: 10.1080/10428194.2018.1516039.
264. Kim, S. *et al.* (2014) 'Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.', *Genome research*. Cold Spring Harbor Laboratory Press, 24(6), pp. 1012–9. doi: 10.1101/gr.171322.113.
265. Kim, Y. B. *et al.* (2017) 'Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions', *Nature Biotechnology*, 35(4), pp. 371–376. doi: 10.1038/nbt.3803.
266. Kingston, R. E., Chen, C. A. and Rose, J. K. (2003) 'Calcium Phosphate Transfection', *Current Protocols in Molecular Biology*. Wiley, 63(1), pp. 9.1.1-9.1.11. doi: 10.1002/0471142727.mb0901s63.
267. Kirstetter, P. *et al.* (2008) 'Modeling of C/EBPα Mutant Acute Myeloid Leukemia Reveals a Common Expression Signature of Committed Myeloid Leukemia-Initiating Cells', *Cancer Cell*. Elsevier, 13(4), pp. 299–310. doi: 10.1016/j.ccr.2008.02.008.
268. Kitajima, K. *et al.* (2002) 'GATA-2 and GATA-2/ER display opposing activities in the development and differentiation of blood progenitors', *EMBO Journal*, 21(12), pp. 3060–3069. doi: 10.1093/emboj/cdf301.
269. Ko, L. J. and Engel, J. D. (1993) 'DNA-binding specificities of the GATA transcription factor family.', *Molecular and cellular biology*. American Society for Microbiology Journals, 13(7), pp. 4011–22. doi: 10.1128/MCB.13.7.4011.
270. Kobayashi, H. *et al.* (2010) 'Angiocrine factors from Akt-activated endothelial cells balance self-renewal and differentiation of haematopoietic stem cells', *Nature Cell Biology*. Nature Publishing Group, 12(11), pp. 1046–1056. doi: 10.1038/ncb2108.
271. Koblan, L. W. *et al.* (2018) 'Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction', *Nature Biotechnology*, 36(9), pp. 843–848. doi: 10.1038/nbt.4172.
272. Koegel, A. K. *et al.* (2016) 'Acute lymphoblastic leukemia in a patient with MonoMAC syndrome/GATA2 haploinsufficiency', *Pediatric Blood and Cancer*. John Wiley and Sons Inc., 63(10), pp. 1844–1847. doi: 10.1002/pbc.26084.
273. Koga, S. *et al.* (2007) 'Cell-cycle-dependent oscillation of GATA2 expression in hematopoietic cells', *Blood*. American Society of Hematology, 109(10), pp. 4200–4208. doi: 10.1182/blood-2006-08-044149.
274. Komor, A. C. *et al.* (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature*, 533, pp. 420–424. doi: 10.1038/nature17946.
275. Komor, A. C., Badran, A. H. and Liu, D. R. (2017) 'CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes', *Cell*. Cell Press, pp. 20–36. doi: 10.1016/j.cell.2016.10.044.
276. Komor, A. C., Badran, A. H. and Liu, D. R. (2018) 'Editing the Genome Without Double-Stranded DNA Breaks', *ACS Chemical Biology*. UTC, pp. 383–388. doi: 10.1021/acschembio.7b00710.
277. Konstandin, N. P. *et al.* (2018) 'Genetic heterogeneity of cytogenetically normal AML with mutations of CEBPA', *Blood Advances*. American Society of Hematology, 2(20), pp. 2724–2731. doi: 10.1182/bloodadvances.2018016840.
278. Korkmaz, G. *et al.* (2016) 'Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9', *Nature Biotechnology*. Nature Publishing Group, 34(2), pp. 192–198. doi: 10.1038/nbt.3450.
279. Korneev, K. V. *et al.* (2017) 'TLR-signaling and proinflammatory cytokines as drivers of tumorigenesis', *Cytokine*. Academic Press, pp. 127–135. doi: 10.1016/j.cyto.2016.01.021.
280. Kotini, A. G. *et al.* (2015) 'Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells', *Nature Biotechnology*. NIH Public Access, 33(6), pp. 646–655. doi: 10.1038/nbt.3178.
281. Kotini, A. G. *et al.* (2017) 'Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia', *Cell Stem Cell*. Cell Press, 20(3), pp. 315–328. doi: 10.1016/j.stem.2017.01.009.
282. Kotini, A. G. and Papapetrou, E. P. (2020) 'Engineering of targeted megabase-scale deletions in human induced pluripotent stem cells', *Experimental Hematology*. Elsevier Inc., 87, pp. 25–32. doi: 10.1016/j.exphem.2020.06.001.
283. De Kouchkovsky, I and Abdul-Hay, M. (2016) '"Acute myeloid leukemia: A comprehensive review and 2016 update"', *Blood Cancer Journal*. doi: 10.1038/bcj.2016.50.
284. De Kouchkovsky, I. and Abdul-Hay, M. (2016) '"Acute myeloid leukemia: A comprehensive review and 2016 update"', *Blood Cancer Journal*. Nature Publishing Group, p. e441. doi: 10.1038/bcj.2016.50.
285. Kouros-Mehr, H. *et al.* (2006) 'GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland', *Cell*. NIH Public Access, 127(5), pp. 1041–1055. doi: 10.1016/j.cell.2006.09.048.

286. Kozyra, E. J. *et al.* (2020) 'Synonymous GATA2 mutations result in selective loss of mutated RNA and are common in patients with GATA2 deficiency', *Leukemia*. Springer Nature, 34(10), pp. 2673–2687. doi: 10.1038/s41375-020-0899-5.
287. Kranc, K. R. *et al.* (2009) 'Cited2 Is an Essential Regulator of Adult Hematopoietic Stem Cells', *Cell Stem Cell*. Elsevier, 5(6), pp. 659–665. doi: 10.1016/j.stem.2009.11.001.
288. Krause, M. N., Sancho-Martinez, I. and Izpisua Belmonte, J. C. (2016) 'Understanding the molecular mechanisms of reprogramming', *Biochemical and Biophysical Research Communications*. Elsevier B.V., pp. 693–697. doi: 10.1016/j.bbrc.2015.11.120.
289. Kricun, M. E. (1985) 'Red-yellow marrow conversion: Its effect on the location of some solitary bone lesions', *Skeletal Radiology*. Springer-Verlag, 14(1), pp. 10–19. doi: 10.1007/BF00361188.
290. Krueger, M. J. *et al.* (2015) 'A Novel STAT3 Inhibitor Has Potent Activity in Preclinical Models of Acute Myeloid Leukemia That Incorporate the Stromal Environment', *Blood*. American Society of Hematology, 126(23), pp. 569–569. doi: 10.1182/blood.v126.23.569.569.
291. Kumar, A. *et al.* (2020) 'Generation of T cells from Human and Nonhuman Primate Pluripotent Stem Cells', *BIO-PROTOCOL*. Bio-Protocol, LLC, 10(13). doi: 10.21769/bioprotoc.3675.
292. Kumar, A., D'Souza, S. S. and Thakur, A. S. (2019) 'Understanding the Journey of Human Hematopoietic Stem Cell Development', *Stem Cells International*. Hindawi, 2019, pp. 1–13. doi: 10.1155/2019/2141475.
293. Kunimoto, H. and Nakajima, H. (2021) 'TET2: A cornerstone in normal and malignant hematopoiesis', *Cancer Science*. Blackwell Publishing Ltd, 112(1), pp. 31–40. doi: 10.1111/cas.14688.
294. Kunisaki, Y. *et al.* (2013) 'Arteriolar niches maintain haematopoietic stem cell quiescence', *Nature*, 502(7473), pp. 637–643. doi: 10.1038/nature12612.
295. Kuvardina, O. N. *et al.* (2015) 'RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation', *Blood*. American Society of Hematology, 125(23), pp. 3570–3579. doi: 10.1182/blood-2014-11-610519.
296. Larochelle, A. *et al.* (1996) 'Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: Implications for gene therapy', *Nature Medicine*. Nature Publishing Group, 2(12), pp. 1329–1337. doi: 10.1038/nm1296-1329.
297. Latger-Cannard, V. *et al.* (2016) 'Haematological spectrum and genotype-phenotype correlations in nine unrelated families with RUNX1 mutations from the French network on inherited platelet disorders', *Orphanet Journal of Rare Diseases*. BioMed Central Ltd., 11(1). doi: 10.1186/s13023-016-0432-0.
298. Lazarus, H. M. and Rowe, J. M. (2006) 'Reduced-intensity conditioning for acute myeloid leukemia: Is this strategy correct', *Leukemia*. Nature Publishing Group, pp. 1673–1682. doi: 10.1038/sj.leu.2404328.
299. Ledran, M. H. *et al.* (2008) 'Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches', *Cell Stem Cell*. Cell Stem Cell, 3(1), pp. 85–98. doi: 10.1016/j.stem.2008.06.001.
300. Lee, G. R., Fields, P. E. and Flavell, R. A. (2001) 'Regulation of IL-4 gene expression by distal regulatory elements and GATA-3 at the chromatin level', *Immunity*. Cell Press, 14(4), pp. 447–459. doi: 10.1016/S1074-7613(01)00125-X.
301. Lee, H. K. *et al.* (2006) 'Double-stranded RNA-mediated TLR3 activation is enhanced by CD14', *Immunity*. Cell Press, 24(2), pp. 153–163. doi: 10.1016/j.immuni.2005.12.012.
302. Lee, J. H. *et al.* (2017) 'Brief Report: Human Acute Myeloid Leukemia Reprogramming to Pluripotency Is a Rare Event and Selects for Patient Hematopoietic Cells Devoid of Leukemic Mutations', *Stem Cells*. Wiley-Blackwell, 35(9), pp. 2095–2102. doi: 10.1002/stem.2655.
303. Leuschner, F. *et al.* (2012) 'Rapid monocyte kinetics in acute myocardial infarction are sustained by extramedullary monocytopoiesis', *Journal of Experimental Medicine*. The Rockefeller University Press, 209(1), pp. 123–137. doi: 10.1084/jem.20111009.
304. Ley, T. J. *et al.* (2013) 'Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia', *New England Journal of Medicine*. Massachussetts Medical Society, 368(22), pp. 2059–2074. doi: 10.1056/NEJMoa1301689.
305. Li, H. S. *et al.* (2016) 'Loss of c-Kit and bone marrow failure upon conditional removal of the GATA-2 C-terminal zinc finger domain in adult mice', *European Journal of Haematology*. Blackwell Publishing Ltd, 97(3), pp. 261–270. doi: 10.1111/ejh.12719.
306. Li, W. and Ding, S. (2010) 'Small molecules that modulate embryonic stem cell fate and somatic cell reprogramming', *Trends in Pharmacological Sciences*, 31(1), pp. 36–45. doi: 10.1016/j.tips.2009.10.002.
307. Li, X. *et al.* (2016) 'Gata2 is a rheostat for mesenchymal stem cell fate in male mice', *Endocrinology*. Endocrine Society, 157(3), pp. 1021–1028. doi: 10.1210/en.2015-1827.

308. Li, Y. *et al.* (2015) 'The STAT5–GATA2 Pathway Is Critical in Basophil and Mast Cell Differentiation and Maintenance', *The Journal of Immunology*. The American Association of Immunologists, 194(9), pp. 4328–4338. doi: 10.4049/jimmunol.1500018.
309. Liang, P. *et al.* (2017) 'Correction of β-thalassemia mutant by base editor in human embryos', *Protein and Cell*. Higher Education Press, 8(11), pp. 811–822. doi: 10.1007/s13238-017-0475-6.
310. Liang, X. *et al.* (2015) 'Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection', *Journal of Biotechnology*, 208, pp. 44–53. doi: 10.1016/j.jbiotec.2015.04.024.
311. van Lier, Y. F. *et al.* (2020) 'Allogeneic hematopoietic cell transplantation in the management of GATA2 deficiency and pulmonary alveolar proteinosis', *Clinical Immunology*. Academic Press Inc., 218, p. 108522. doi: 10.1016/j.clim.2020.108522.
312. Lim, K. C. *et al.* (2012) 'Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning', *Journal of Clinical Investigation*, 122(10), pp. 3705–3717. doi: 10.1172/JCI61619.
313. Lin, S. *et al.* (2014) 'Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.', *eLife*. eLife Sciences Publications, Ltd, 3, p. e04766. doi: 10.7554/eLife.04766.
314. Lindsley, R. C. and Ebert, B. L. (2013) 'Molecular pathophysiology of myelodysplastic syndromes', *Annual Review of Pathology: Mechanisms of Disease*. Annual Reviews Inc., pp. 21–47. doi: 10.1146/annurev-pathol-011811-132436.
315. Ling, K.-W. W. *et al.* (2004) 'GATA-2 Plays Two Functionally Distinct Roles during the Ontogeny of Hematopoietic Stem Cells', *The Journal of Experimental Medicine*. The Rockefeller University Press, 200(7), pp. 871–882. doi: 10.1084/jem.20031556.
316. Ling, K. W. *et al.* (2004) 'GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells', *Journal of Experimental Medicine*. J Exp Med, 200(7), pp. 871–882. doi: 10.1084/jem.20031556.
317. Lio, C. W. J., Yuita, H. and Rao, A. (2019) 'Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies', *Blood*. American Society of Hematology, pp. 1487–1497. doi: 10.1182/blood.2019791475.
318. Liu, G. H. *et al.* (2014) 'Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs', *Nature Communications*. Nature Publishing Group, 5, p. 4330. doi: 10.1038/ncomms5330.
319. Liu, M. *et al.* (2018) 'Methodologies for Improving HDR Efficiency.', *Frontiers in genetics*. Frontiers Media SA, 9, p. 691. doi: 10.3389/fgene.2018.00691.
320. Liu, Z. *et al.* (2018) 'Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing', *Nature Communications*. Nature Publishing Group, 9(1), pp. 1–8. doi: 10.1038/s41467-018-04768-7.
321. Loberg, M. A. *et al.* (2019) 'Sequentially inducible mouse models reveal that Npm1 mutation causes malignant transformation of Dnmt3a-mutant clonal hematopoiesis', *Leukemia*. Nature Publishing Group, 33(7), pp. 1635–1649. doi: 10.1038/s41375-018-0368-6.
322. Lu, R. (2014) 'Sleeping beauty wakes up the clonal succession model for homeostatic hematopoiesis', *Cell Stem Cell*. Cell Press, pp. 677–678. doi: 10.1016/j.stem.2014.11.015.
323. Lübking, A. *et al.* (2015) 'Young woman with mild bone marrow dysplasia, GATA2 and ASXL1 mutation treated with allogeneic hematopoietic stem cell transplantation', *Leukemia Research Reports*. Elsevier Ltd, 4(2), pp. 72–75. doi: 10.1016/j.lrr.2015.10.001.
324. Ludin, A. *et al.* (2012) 'Monocytes-macrophages that express α-smooth muscle actin preserve primitive hematopoietic cells in the bone marrow', *Nature Immunology*. Nature Publishing Group, 13(11), pp. 1072–1082. doi: 10.1038/ni.2408.
325. Luesink, M. *et al.* (2012) 'High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia', *Blood*, 120(10), pp. 2064–2075. doi: 10.1182/blood-2011-12-397083.
326. Luo, X. *et al.* (2019) 'ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants', *Blood Advances*. American Society of Hematology, 3(20), pp. 2962–2979. doi: 10.1182/bloodadvances.2019000644.
327. Macaulay, I. C. *et al.* (2016) 'Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells', *Cell Reports*. Elsevier B.V., 14(4), pp. 966–977. doi: 10.1016/j.celrep.2015.12.082.
328. Mace, E. M. *et al.* (2013) 'Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56bright subset', *Blood*. American Society of Hematology, 121(14), pp. 2669–2677. doi: 10.1182/blood-2012-09-453969.
329. Maciejewski, J. P. *et al.* (2017) 'DDX41-related myeloid neoplasia', *Seminars in Hematology*. W.B. Saunders, pp. 94–97. doi: 10.1053/j.seminhematol.2017.04.007.

330. Maeurer, M. *et al.* (2014) 'Allogeneic hematopoietic cell transplantation for GATA2 deficiency in a patient with disseminated human papillomavirus disease', *Transplantation*. Lippincott Williams and Wilkins, pp. e95–e96. doi: 10.1097/TP.0000000000000520.
331. Mallhi, K. *et al.* (2016) 'Successful umbilical cord blood hematopoietic stem cell transplantation in pediatric patients with MDS/AML associated with underlying GATA2 mutations: two case reports and review of literature', *Pediatric Transplantation*, 20(7), pp. 1004–1007. doi: 10.1111/petr.12764.
332. Månsson, R. *et al.* (2007) 'Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors', *Immunity*. Cell Press, 26(4), pp. 407–419. doi: 10.1016/j.immuni.2007.02.013.
333. Matheny, C. J. *et al.* (2007) 'Disease mutations in RUNX1 and RUNX2 create nonfunctional, dominant-negative, or hypomorphic alleles', *EMBO Journal*. EMBO J, 26(4), pp. 1163–1175. doi: 10.1038/sj.emboj.7601568.
334. Matos, A., Magalhães, S. M. M. and Rauh, M. J. (2020) 'Immune Dysregulation and Recurring Mutations in Myelodysplastic Syndromes Pathogenesis', in. Springer, New York, NY, pp. 1–10. doi: 10.1007/5584_2020_608.
335. McBride, J. A. and Striker, R. (2017) 'Imbalance in the game of T cells: What can the CD4/CD8 T-cell ratio tell us about HIV and health?', *PLoS Pathogens*. Public Library of Science, 13(11). doi: 10.1371/journal.ppat.1006624.
336. McGrath, E. *et al.* (2019) 'Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing.', *Nature communications*, 10(1), p. 5353. doi: 10.1038/s41467-019-13342-8.
337. McGrath, K. E. *et al.* (2015) 'Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo', *Cell Reports*. The Authors, 11(12), pp. 1892–1904. doi: 10.1016/j.celrep.2015.05.036.
338. McNerney, M. E. *et al.* (2013) 'CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia', *Blood*. The American Society of Hematology, 121(6), pp. 975–983. doi: 10.1182/blood-2012-04-426965.
339. McReynolds, L. J., Yang, Y., *et al.* (2019) 'MDS-associated mutations in germline GATA2 mutated patients with hematologic manifestations', *Leukemia Research*. Pergamon, 76, pp. 70–75. doi: 10.1016/J.LEUKRES.2018.11.013.
340. McReynolds, L. J., Zhang, Y., *et al.* (2019) 'Rapid progression to AML in a patient with germline GATA2 mutation and acquired NRAS Q61K mutation', *Leukemia Research Reports*. Elsevier Ltd, 12. doi: 10.1016/j.lrr.2019.100176.
341. McReynolds, L. J., Calvo, K. R. and Holland, S. M. (2018) 'Germline GATA2 Mutation and Bone Marrow Failure', *Hematology/Oncology Clinics of North America*, pp. 713–728. doi: 10.1016/j.hoc.2018.04.004.
342. Mehta, C. *et al.* (2017) 'Integrating Enhancer Mechanisms to Establish a Hierarchical Blood Development Program', *Cell Reports*. Elsevier B.V., 20(12), pp. 2966–2979. doi: 10.1016/j.celrep.2017.08.090.
343. Melazzini, F. *et al.* (2016) 'Clinical and pathogenic features of ETV6-related thrombocytopenia with predisposition to acute lymphoblastic leukemia', *Haematologica*. Ferrata Storti Foundation, 101(11), pp. 1333–1342. doi: 10.3324/haematol.2016.147496.
344. Mendes-De-Almeida, D. P. *et al.* (2019) 'GATA2 mutation in long stand Mycobacterium kansasii infection, myelodysplasia and MonoMAC syndrome: A case-report', *BMC Medical Genetics*. BioMed Central Ltd., 20(1), p. 64. doi: 10.1186/s12881-019-0799-6.
345. Menendez-Gonzalez, J. B. *et al.* (2019) 'Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia', *Stem Cell Reports*. Cell Press, 13(2), pp. 291–306. doi: 10.1016/j.stemcr.2019.07.005.
346. Merika, M. and Orkin, S. H. (1993) 'DNA-binding specificity of GATA family transcription factors.', *Molecular and Cellular Biology*. American Society for Microbiology, 13(7), pp. 3999–4010. doi: 10.1128/mcb.13.7.3999.
347. Michaud, J. *et al.* (2002) 'In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: Implications for mechanisms of pathogenesis', *Blood*. American Society of Hematology, 99(4), pp. 1364–1372. doi: 10.1182/blood.V99.4.1364.
348. Mikkola, H. K. A. *et al.* (2003) 'Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene', *Nature*. Nature Publishing Group, 421(6922), pp. 547–551. doi: 10.1038/nature01345.
349. Minegishi, N. *et al.* (1998) 'Alternative promoters regulate transcription of the mouse GATA-2 gene',

Journal of Biological Chemistry. American Society for Biochemistry and Molecular Biology, 273(6), pp. 3625–3634. doi: 10.1074/jbc.273.6.3625.
350. Minegishi, N. *et al.* (2003) 'Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos', *Blood*, 102(3), pp. 896–905. doi: 10.1182/blood-2002-12-3809.
351. Minehata, K. I. *et al.* (2006) 'Oncostatin M maintains the hematopoietic microenvironment and retains hematopoietic progenitors in the bone marrow', *International Journal of Hematology*. Int J Hematol, 84(4), pp. 319–327. doi: 10.1532/IJH97.06090.
352. Mitchell, K. *et al.* (2018) 'IL1RAP potentiates multiple oncogenic signaling pathways in AML', *Journal of Experimental Medicine*. Rockefeller University Press, 215(6), pp. 1709–1727. doi: 10.1084/jem.20180147.
353. Mitroulis, I. *et al.* (2020) 'Regulation of the Bone Marrow Niche by Inflammation', *Frontiers in Immunology*. Frontiers Media S.A., p. 1540. doi: 10.3389/fimmu.2020.01540.
354. Montagner, S. *et al.* (2016) 'TET2 Regulates Mast Cell Differentiation and Proliferation through Catalytic and Non-catalytic Activities', *Cell Reports*. Elsevier B.V., 15(7), pp. 1566–1579. doi: 10.1016/j.celrep.2016.04.044.
355. Montalbano, A., Canver, M. C. and Sanjana, N. E. (2017) 'High-Throughput Approaches to Pinpoint Function within the Noncoding Genome', *Molecular Cell*. Cell Press, pp. 44–59. doi: 10.1016/j.molcel.2017.09.017.
356. Moralli, D. *et al.* (2011) 'An Improved Technique for Chromosomal Analysis of Human ES and iPS Cells', *Stem Cell Reviews and Reports*, 7(2), pp. 471–477. doi: 10.1007/s12015-010-9224-4.
357. Moriguchi, T. and Yamamoto, M. (2014) 'A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation', *International journal of hematology*, 100(5), pp. 417–424. doi: 10.1007/s12185-014-1568-0.
358. Morrisey, E. E. *et al.* (1997) *GATA-4 Activates Transcription Via Two Novel Domains That Are Conserved within the GATA-4/5/6 Subfamily**. Available at: http://www-jbc.stanford.edu/jbc/ (Accessed: 22 May 2020).
359. Morrison, S. J. *et al.* (1997) 'Identification of a lineage of multipotent hematopoietic progenitors', *Development*, 124(10), pp. 1929–1939.
360. Morrison, S. J. and Scadden, D. T. (2014) 'The bone marrow niche for haematopoietic stem cells.', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 505(7483), pp. 327–34. doi: 10.1038/nature12984.
361. Morrison, S. J. and Weissman, I. L. (1994) 'The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype', *Immunity*. Cell Press, 1(8), pp. 661–673. doi: 10.1016/1074-7613(94)90037-X.
362. Müller, L. U. W. *et al.* (2012) 'Overcoming reprogramming resistance of Fanconi anemia cells', *Blood*. Blood, 119(23), pp. 5449–5457. doi: 10.1182/blood-2012-02-408674.
363. Muñoz-López, A. *et al.* (2016) 'Development Refractoriness of MLL-Rearranged Human B Cell Acute Leukemias to Reprogramming into Pluripotency', *Stem Cell Reports*. Cell Press, 7(4), pp. 602–618. doi: 10.1016/j.stemcr.2016.08.013.
364. Muto, T. *et al.* (2020) 'Adaptive response to inflammation contributes to sustained myelopoiesis and confers a competitive advantage in myelodysplastic syndrome HSCs', *Nature Immunology*. Nature Research, 21(5), pp. 535–545. doi: 10.1038/s41590-020-0663-z.
365. Na Nakorn, T. *et al.* (2002) 'Myeloerythroid-restricted progenitors are sufficient to confer radioprotection and provide the majority of day 8 CFU-S', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 109(12), pp. 1579–1585. doi: 10.1172/jci15272.
366. Nagamachi, A. *et al.* (2013) 'Haploinsufficiency of SAMD9L, an Endosome Fusion Facilitator, Causes Myeloid Malignancies in Mice Mimicking Human Diseases with Monosomy 7', *Cancer Cell*. Cancer Cell, 24(3), pp. 305–317. doi: 10.1016/j.ccr.2013.08.011.
367. Nandakumar, S. K. *et al.* (2015) 'Low-level GATA2 overexpression promotes myeloid progenitor self-renewal and blocks lymphoid differentiation in mice', *Experimental Hematology*. Elsevier Ltd, 43(7), pp. 565-577.e10. doi: 10.1016/j.exphem.2015.04.002.
368. Narayanan, D. and Weinberg, O. K. (2020) 'How I investigate acute myeloid leukemia', *International Journal of Laboratory Hematology*. Blackwell Publishing Ltd, pp. 3–15. doi: 10.1111/ijlh.13135.
369. Navarro, S. *et al.* (2017) 'Induced Pluripotency and Gene Editing in Fanconi Anemia', *Current Gene Therapy*. Bentham Science Publishers Ltd., 16(5), pp. 321–328. doi: 10.2174/1566523217666170118112050.
370. Naveiras, O. *et al.* (2009) 'Bone-marrow adipocytes as negative regulators of the haematopoietic

microenvironment', *Nature*. Nature Publishing Group, 460(7252), pp. 259–263. doi: 10.1038/nature08099.
371. Nawijn, M. C. *et al.* (2001) 'Enforced Expression of GATA-3 During T Cell Development Inhibits Maturation of CD8 Single-Positive Cells and Induces Thymic Lymphoma in Transgenic Mice', *The Journal of Immunology*. The American Association of Immunologists, 167(2), pp. 715–723. doi: 10.4049/jimmunol.167.2.715.
372. Nei, Y. *et al.* (2013) 'GATA-1 regulates the generation and function of basophils', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(46), pp. 18620–18625. doi: 10.1073/pnas.1311668110.
373. Ng, E. S. *et al.* (2005) 'Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation', *Blood*, 106(5), pp. 1601–1603. doi: 10.1182/blood-2005-03-0987.
374. Nick, H. J. *et al.* (2012) 'Distinct classes of c-Kit-activating mutations differ in their ability to promote RUNX1-ETO-associated acute myeloid leukemia', *Blood*. Blood, 119(6), pp. 1522–1531. doi: 10.1182/blood-2011-02-338228.
375. Noguchi, H., Miyagi-Shiohira, C. and Nakashima, Y. (2018) 'Induced tissue-specific stem cells and epigenetic memory in induced pluripotent stem cells', *International Journal of Molecular Sciences*. MDPI AG, p. 930. doi: 10.3390/ijms19040930.
376. Nombela-Arrieta, C. *et al.* (2013) 'Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment', *Nature Cell Biology*. Nat Cell Biol, 15(5), pp. 533–543. doi: 10.1038/ncb2730.
377. Notta, F. *et al.* (2011) 'Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment', *Science*. Science, 333(6039), pp. 218–221. doi: 10.1126/science.1201219.
378. Notta, F. *et al.* (2016) 'Distinct routes of lineage development reshape the human blood hierarchy across ontogeny', *Science*. American Association for the Advancement of Science, 351(6269). doi: 10.1126/science.aab2116.
379. Oguro, H., Ding, L. and Morrison, S. J. (2013) 'SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors', *Cell Stem Cell*. Cell Press, 13(1), pp. 102–116. doi: 10.1016/j.stem.2013.05.014.
380. Ohmori, S. *et al.* (2019) 'GATA2 and PU.1 Collaborate To Activate the Expression of the Mouse Ms4a2 Gene, Encoding FcεRIβ, through Distinct Mechanisms', *Molecular and Cellular Biology*. American Society for Microbiology, 39(22), pp. e00314-19. doi: 10.1128/mcb.00314-19.
381. Ohmori, S. N. Y. *et al.* (2015) 'GATA2 is critical for the maintenance of cellular identity in differentiated mast cells derived from mouse bone marrow', *Blood*. American Society of Hematology, 125(21), pp. 3306–3315. doi: 10.1182/blood-2014-11-612465.
382. Ohneda, K. *et al.* (2014) 'Transcription Factor GATA1 Is Dispensable for Mast Cell Differentiation in Adult Mice', *Molecular and Cellular Biology*. American Society for Microbiology, 34(10), pp. 1812–1826. doi: 10.1128/mcb.01524-13.
383. Okita, K., Ichisaka, T. and Yamanaka, S. (2007) 'Generation of germline-competent induced pluripotent stem cells.', *Nature*, 448(7151), pp. 313–317. doi: 10.1038/nature05934.
384. Oleaga-Quintas, C. *et al.* (2021) 'Inherited GATA2 Deficiency Is Dominant by Haploinsufficiency and Displays Incomplete Clinical Penetrance', *Journal of Clinical Immunology*. Springer. doi: 10.1007/s10875-020-00930-3.
385. Omatsu, Y. *et al.* (2010) 'The Essential Functions of Adipo-osteogenic Progenitors as the Hematopoietic Stem and Progenitor Cell Niche', *Immunity*. Cell Press, 33(3), pp. 387–399. doi: 10.1016/j.immuni.2010.08.017.
386. Onder, T. T. *et al.* (2012) 'Chromatin-modifying enzymes as modulators of reprogramming', *Nature*. Nature Publishing Group, 483(7391), pp. 598–602. doi: 10.1038/nature10953.
387. Onodera, K. *et al.* (2016) 'GATA2 regulates dendritic cell differentiation', *Blood*. American Society of Hematology, 128(4), pp. 508–518. doi: 10.1182/blood-2016-02-698118.
388. Orkin, S. H. (1995) 'Transcription factors and hematopoietic development.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 270(10), pp. 4955–8. doi: 10.1074/JBC.270.10.4955.
389. Orkin, S. H. (2000) 'Diversification of haematopoietic stem cells to specific lineages', *Nature Reviews Genetics*, 1(1), pp. 57–64. doi: 10.1038/35049577.
390. Orkin, S. H. and Zon, L. I. (2008) 'Hematopoiesis: an evolving paradigm for stem cell biology.', *Cell*. Elsevier, 132(4), pp. 631–44. doi: 10.1016/j.cell.2008.01.025.
391. Ortmann, C. A. *et al.* (2015) 'Effect of Mutation Order on Myeloproliferative Neoplasms', *New England*

Journal of Medicine. New England Journal of Medicine (NEJM/MMS), 372(7), pp. 601–612. doi: 10.1056/nejmoa1412098.
392. Ostergaard, P. *et al.* (2011) 'Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome)', *Nature Genetics*, 43(10), pp. 929–931. doi: 10.1038/ng.923.
393. Ottersbach, K. and Dzierzak, E. (2005) 'The murine placenta contains hematopoietic stem cells within the vascular labyrinth region', *Developmental Cell*. Dev Cell, 8(3), pp. 377–387. doi: 10.1016/j.devcel.2005.02.001.
394. van Overbeek, M. *et al.* (2016) 'DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks', *Molecular Cell*. Cell Press, 63(4), pp. 633–646. doi: 10.1016/j.molcel.2016.06.037.
395. Owen, C. J. *et al.* (2008) 'Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy', *Blood*. American Society of Hematology, 112(12), pp. 4639–4645. doi: 10.1182/blood-2008-05-156745.
396. Pabst, T. *et al.* (2008) 'Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 26(31), pp. 5088–5093. doi: 10.1200/JCO.2008.16.5563.
397. Pajerowski, A. G. *et al.* (2010) 'Adult hematopoietic stem cells require NKAP for maintenance and survival', *Blood*. American Society of Hematology, 116(15), pp. 2684–2693. doi: 10.1182/blood-2010-02-268391.
398. Pan, X. *et al.* (2000) 'Identification of human GATA-2 gene distal IS exon and its expression in hematopoietic stem cell fractions', *Journal of Biochemistry*. Japanese Biochemical Society, 127(1), pp. 105–112. doi: 10.1093/oxfordjournals.jbchem.a022570.
399. Papaemmanuil, E. *et al.* (2013) 'Clinical and biological implications of driver mutations in myelodysplastic syndromes', *Blood*. Blood, 122(22), pp. 3616–3627. doi: 10.1182/blood-2013-08-518886.
400. Pasque, V. and Plath, K. (2015) 'X chromosome reactivation in reprogramming and in development', *Current Opinion in Cell Biology*. Elsevier Ltd, pp. 75–83. doi: 10.1016/j.ceb.2015.10.006.
401. de Pater, E. *et al.* (2013) 'Gata2 is required for HSC generation and survival', *Journal of Experimental Medicine*, 210(13), pp. 2843–2850. doi: 10.1084/jem.20130751.
402. Patient, R. K. and McGhee, J. D. (2002) 'The GATA family (vertebrates and invertebrates)', *Current Opinion in Genetics & Development*. Elsevier Current Trends, 12(4), pp. 416–422. doi: 10.1016/S0959-437X(02)00319-2.
403. Paulsson, K. *et al.* (2001) 'The incidence of trisomy 8 as a sole chromosomal aberration in myeloid malignancies varies in relation to gender, age, prior iatrogenic genotoxic exposure, and morphology', *Cancer Genetics and Cytogenetics*. Elsevier, 130(2), pp. 160–165. doi: 10.1016/S0165-4608(01)00486-1.
404. Paz-Priel, I. *et al.* (2011) 'C/EBPA and C/EBPA oncoproteins regulate nfkb1 and displace histone deacetylases from NF-κB p50 homodimers to induce NF-κB target genes', *Blood*. The American Society of Hematology, 117(15), pp. 4085–4094. doi: 10.1182/blood-2010-07-294470.
405. Peignier, A. and Parker, D. (2021) 'Impact of Type I Interferons on Susceptibility to Bacterial Pathogens', *Trends in Microbiology*. Elsevier Ltd. doi: 10.1016/j.tim.2021.01.007.
406. Pellagatti, A. *et al.* (2018) 'Impact of spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated genes/pathways and clinical associations', *Blood*. American Society of Hematology, 132(12), pp. 1225–1240. doi: 10.1182/blood-2018-04-843771.
407. Perez-Andreu, V. *et al.* (2013) 'Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse', *Nature Genetics*. NIH Public Access, 45(12), pp. 1494–1498. doi: 10.1038/ng.2803.
408. Pérez-Figueroa, E. *et al.* (2021) 'Neutrophils: Many Ways to Die', *Frontiers in Immunology*. Frontiers Media S.A. doi: 10.3389/fimmu.2021.631821.
409. Persons, D. and Allay, J. (1999) 'Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis', *Blood*, 93(2), pp. 488–99. Available at: http://bloodjournal.hematologylibrary.org/content/93/2/488.short.
410. Pevny, L. *et al.* (1991) 'Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1', *Nature*. Nature Publishing Group, 349(6306), pp. 257–260. doi: 10.1038/349257a0.
411. Pevny, L. *et al.* (1995) 'Development of hematopoietic cells lacking transcription factor GATA-1', *Development*, 121(1), pp. 163–172.
412. Pick, M. *et al.* (2007) 'Differentiation of Human Embryonic Stem Cells in Serum-Free Medium Reveals

Distinct Roles for Bone Morphogenetic Protein 4, Vascular Endothelial Growth Factor, Stem Cell Factor, and Fibroblast Growth Factor 2 in Hematopoiesis', *STEM CELLS*. Wiley, 25(9), pp. 2206–2214. doi: 10.1634/stemcells.2006-0713.
413. Pickar-Oliver, A. and Gersbach, C. A. (2019) 'The next generation of CRISPR–Cas technologies and applications', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 490–507. doi: 10.1038/s41580-019-0131-5.
414. Pietras, E. M. *et al.* (2014) 'Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferons', *Journal of Experimental Medicine*. J Exp Med, 211(2), pp. 245–262. doi: 10.1084/jem.20131043.
415. Pietras, E. M. *et al.* (2015) 'Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions', *Cell Stem Cell*. Cell Press, 17(1), pp. 35–46. doi: 10.1016/j.stem.2015.05.003.
416. Pietras, E. M. (2017) 'Inflammation: A key regulator of hematopoietic stem cell fate in health and disease', *Blood*. American Society of Hematology, 130(15), pp. 1693–1698. doi: 10.1182/blood-2017-06-780882.
417. Pietras, E. M., Warr, M. R. and Passegué, E. (2011) 'Cell cycle regulation in hematopoietic stem cells', *Journal of Cell Biology*. The Rockefeller University Press, pp. 709–720. doi: 10.1083/jcb.201102131.
418. Ping, N. *et al.* (2017) 'Exome sequencing identifies highly recurrent somatic GATA2 and CEBPA mutations in acute erythroid leukemia', *Leukemia*. Nature Publishing Group, 31(1), pp. 195–202. doi: 10.1038/leu.2016.162.
419. Pinho, S. *et al.* (2018) 'Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches', *Developmental Cell*. Cell Press, 44(5), pp. 634-641.e4. doi: 10.1016/j.devcel.2018.01.016.
420. Pinho, S. and Frenette, P. S. (2019) 'Haematopoietic stem cell activity and interactions with the niche', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 303–320. doi: 10.1038/s41580-019-0103-9.
421. Platzbecker, U. (2019) 'Treatment of MDS', *Blood*. American Society of Hematology, pp. 1096–1107. doi: 10.1182/blood-2018-10-844696.
422. Pollyea, D. A. *et al.* (2019) 'Venetoclax for AML: Changing the treatment paradigm', *Blood Advances*. American Society of Hematology, 3(24), pp. 4326–4335. doi: 10.1182/bloodadvances.2019000937.
423. Poloni, A. *et al.* (2013) 'Molecular and functional characterization of human bone marrow adipocytes', *Experimental Hematology*. Elsevier Inc., 41(6), pp. 558-566.e2. doi: 10.1016/j.exphem.2013.02.005.
424. Polprasert, C. *et al.* (2015) 'Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms', *Cancer Cell*. Cell Press, 27(5), pp. 658–670. doi: 10.1016/j.ccell.2015.03.017.
425. Porcher, C., Chagraoui, H. and Kristiansen, M. S. (2017) 'SCL/TAL1: a multifaceted regulator from blood development to disease.', *Blood*. American Society of Hematology, 129(15), pp. 2051–2060. doi: 10.1182/blood-2016-12-754051.
426. Poulos, M. G. *et al.* (2013) 'Endothelial Jagged-1 Is necessary for homeostatic and regenerative hematopoiesis', *Cell Reports*. NIH Public Access, 4(5), pp. 1022–1034. doi: 10.1016/j.celrep.2013.07.048.
427. Pronk, C. J. H. and Bryder, D. (2018) 'Immunophenotypic identification of early myeloerythroid development', in *Methods in Molecular Biology*. Humana Press Inc., pp. 301–319. doi: 10.1007/978-1-4939-7346-0_13.
428. Pulikkan, J. A., Tenen, D. G. and Behre, G. (2017) 'C/EBPα deregulation as a paradigm for leukemogenesis', *Leukemia*. Nature Publishing Group, pp. 2279–2285. doi: 10.1038/leu.2017.229.
429. Purton, L. E. and Scadden, D. T. (2007) 'Limiting Factors in Murine Hematopoietic Stem Cell Assays', *Cell Stem Cell*, pp. 263–270. doi: 10.1016/j.stem.2007.08.016.
430. Qiu, C. *et al.* (2005) 'Differentiation of human embryonic stem cells into hematopoietic cells by coculture with human fetal liver cells recapitulates the globin switch that occurs early in development', *Experimental Hematology*. Elsevier, 33(12), pp. 1450–1458. doi: 10.1016/j.exphem.2005.09.003.
431. Qiu, P. *et al.* (2004) 'Mutation detection using Surveyor™ nuclease', *BioTechniques*. Eaton Publishing Company, pp. 702–707. doi: 10.2144/04364pf01.
432. Quesada, A. E. *et al.* (2019) 'DDX41 mutations in myeloid neoplasms are associated with male gender, TP53 mutations and high-risk disease', *American Journal of Hematology*. Wiley-Liss Inc., 94(7), pp. 757–766. doi: 10.1002/ajh.25486.
433. Ramaraj, P. *et al.* (2004) 'Effect of mutational inactivation of tyrosine kinase activity on BCR/ABL-induced abnormalities in cell growth and adhesion in human hematopoietic progenitors', *Cancer Research*. American Association for Cancer Research, 64(15), pp. 5322–5331. doi: 10.1158/0008-5472.CAN-03-3656.

434. Ramirez-Alejo, N. and Santos-Argumedo, L. (2014) 'Innate defects of the IL-12/IFN-γ axis in susceptibility to infections by mycobacteria and salmonella', *Journal of Interferon and Cytokine Research*. Mary Ann Liebert Inc., pp. 307–317. doi: 10.1089/jir.2013.0050.
435. Ramos-Mejia, V. *et al.* (2010) 'Nodal/Activin Signaling Predicts Human Pluripotent Stem Cell Lines Prone to Differentiate Toward the Hematopoietic Lineage', *Molecular Therapy*. Cell Press, 18(12), pp. 2173–2181. doi: 10.1038/mt.2010.179.
436. Ramsfjell, V. *et al.* (1999) 'Distinct requirements for optimal growth and In vitro expansion of human CD34(+)CD38(-) bone marrow long-term culture-initiating cells (LTC-IC), extended LTC-IC, and murine in vivo long-term reconstituting stem cells.', *Blood*, 94(12), pp. 4093–102. Available at: http://www.bloodjournal.org/content/bloodjournal/94/12/4093.full.pdf (Accessed: 19 March 2018).
437. Ramzan, M. *et al.* (2017) 'Successful Myeloablative Matched Unrelated Donor Hematopoietic Stem Cell Transplantation in a Young Girl with GATA2 Deficiency and Emberger Syndrome', *Journal of Pediatric Hematology/Oncology*. Lippincott Williams and Wilkins, 39(3), pp. 230–232. doi: 10.1097/MPH.0000000000000737.
438. Real, P. J. *et al.* (2012) 'SCL/TAL1 regulates hematopoietic specification from human embryonic stem cells', in *Molecular Therapy*. Nature Publishing Group, pp. 1443–1453. doi: 10.1038/mt.2012.49.
439. Redell, M. S. *et al.* (2011) 'Stat3 signaling in acute myeloid leukemia: Ligand-dependent and -independent activation and induction of apoptosis by a novel small-molecule Stat3 inhibitor', *Blood*. The American Society of Hematology, 117(21), pp. 5701–5709. doi: 10.1182/blood-2010-04-280123.
440. Rees, H. A. and Liu, D. R. (2018) 'Base editing: precision chemistry on the genome and transcriptome of living cells', *Nature Reviews Genetics*. Nature Publishing Group, pp. 770–788. doi: 10.1038/s41576-018-0059-1.
441. Reimer, J. *et al.* (2017) 'CRISPR-Cas9-induced t(11;19)/MLL-ENL translocations initiate leukemia in human hematopoietic progenitor cells in vivo', *Haematologica*. Ferrata Storti Foundation, 102(9), pp. 1558–1566. doi: 10.3324/haematol.2017.164046.
442. Renneville, A. *et al.* (2008) 'Cooperating gene mutations in acute myeloid leukemia: A review of the literature', *Leukemia*. Nature Publishing Group, pp. 915–931. doi: 10.1038/leu.2008.19.
443. Rhoades, K. L. *et al.* (2000) 'Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model', *Blood*. American Society of Hematology, 96(6), pp. 2108–2115. doi: 10.1182/blood.v96.6.2108.
444. Ribeiro, S. *et al.* (2012) 'Plasmid DNA Size Does Affect Nonviral Gene Delivery Efficiency in Stem Cells', *Cellular Reprogramming*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA , 14(2), pp. 130–137. doi: 10.1089/cell.2011.0093.
445. Rodrigues, N. P. *et al.* (2005) 'Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis', *Blood*, 106(2), pp. 477–484. doi: 10.1182/blood-2004-08-2989.
446. Rodrigues, N. P., Boyd, A. S., Fugazza, C., May, G. E., Guo, YanPing Ping, *et al.* (2008) 'GATA-2 regulates granulocyte-macrophage progenitor cell function', *Blood*. American Society of Hematology, 112(13), pp. 4862–4873. doi: 10.1182/blood-2008-01-136564.
447. Rodrigues, N. P., Boyd, A. S., Fugazza, C., May, G. E., Guo, Yan Ping, *et al.* (2008) 'GATA-2 regulates granulocyte-macrophage progenitor cell function', *Blood*. American Society of Hematology, 112(13), pp. 4862–4873. doi: 10.1182/blood-2008-01-136564.
448. Rodrigues, N. P. *et al.* (2012) 'GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia', *International Journal of Biochemistry and Cell Biology*. Elsevier Ltd, 44(3), pp. 457–460. doi: 10.1016/j.biocel.2011.12.004.
449. Romee, R. *et al.* (2012) 'Cytokine activation induces human memory-like NK cells', *Blood*. The American Society of Hematology, 120(24), pp. 4751–4760. doi: 10.1182/blood-2012-04-419283.
450. Rothenberg, E. V. (2013) 'GATA-3 locks the door to the B-cell option', *Blood*. The American Society of Hematology, pp. 1673–1674. doi: 10.1182/blood-2013-01-477737.
451. Saettini, F. *et al.* (2020) 'When to suspect GATA2 deficiency in pediatric patients: from complete blood count to diagnosis', *Pediatric Hematology and Oncology*. Taylor & Francis, pp. 1–5. doi: 10.1080/08880018.2020.1863536.
452. Sahtoe, D. D. *et al.* (2016) 'BAP1/ASXL1 recruitment and activation for H2A deubiquitination', *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms10292.
453. Sakurai, M. *et al.* (2014) 'Impaired hematopoietic differentiation of RUNX1-mutated induced pluripotent stem cells derived from FPD/AML patients.', *Leukemia*. Nature Publishing Group, 28(October 2013), pp. 1–11. doi: 10.1038/leu.2014.136.
454. Sallman, D. A. and List, A. (2019) 'The central role of inflammatory signaling in the pathogenesis of myelodysplastic syndromes', *Blood*. American Society of Hematology, pp. 1039–1048. doi:

10.1182/blood-2018-10-844654.
455. Sanalkumar, R. *et al.* (2014) 'Mechanism governing a stem cell-generating cis-regulatory element', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(12), p. E1091. doi: 10.1073/pnas.1400065111.
456. Sanchez, R. and Zhou, M. M. (2011) 'The PHD finger: A versatile epigenome reader', *Trends in Biochemical Sciences*. NIH Public Access, p. 364. doi: 10.1016/j.tibs.2011.03.005.
457. Sanda, T. *et al.* (2012) 'Core Transcriptional Regulatory Circuit Controlled by the TAL1 Complex in Human T Cell Acute Lymphoblastic Leukemia', *Cancer Cell*. Elsevier, 22(2), pp. 209–221. doi: 10.1016/j.ccr.2012.06.007.
458. Sano, S. *et al.* (2018) 'Tet2-Mediated Clonal Hematopoiesis Accelerates Heart Failure Through a Mechanism Involving the IL-1β/NLRP3 Inflammasome', *Journal of the American College of Cardiology*. Elsevier USA, 71(8), pp. 875–886. doi: 10.1016/j.jacc.2017.12.037.
459. Sarig, R. *et al.* (2010) 'Mutant p53 facilitates somatic cell reprogramming and augments the malignant potential of reprogrammed cells', *Journal of Experimental Medicine*. The Rockefeller University Press, 207(10), pp. 2127–2140. doi: 10.1084/jem.20100797.
460. Sasaguri, H. *et al.* (2018) 'Introduction of pathogenic mutations into the mouse Psen1 gene by Base Editor and Target-AID', *Nature Communications*. Nature Publishing Group, 9(1). doi: 10.1038/s41467-018-05262-w.
461. Schessl, C. *et al.* (2005) 'The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice', *Journal of Clinical Investigation*. J Clin Invest, 115(8), pp. 2159–2168. doi: 10.1172/JCI24225.
462. Scheuermann, J. C. *et al.* (2010) 'Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB', *Nature*. Nature Publishing Group, 465(7295), pp. 243–247. doi: 10.1038/nature08966.
463. Schlums, H. *et al.* (2017) 'Adaptive NK cells can persist in patients with GATA2 mutation depleted of stem and progenitor cells', *Blood*. American Society of Hematology, 129(14), pp. 1927–1939. doi: 10.1182/blood-2016-08-734236.
464. Schmidt, L., Heyes, E. and Grebien, F. (2020) 'Gain-of-Function Effects of N-Terminal CEBPA Mutations in Acute Myeloid Leukemia', *BioEssays*. John Wiley and Sons Inc., 42(2), p. 1900178. doi: 10.1002/bies.201900178.
465. Schmitt, T. M. and Zúñiga-Pflücker, J. C. (2002) 'Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro', *Immunity*. Cell Press, 17(6), pp. 749–756. doi: 10.1016/S1074-7613(02)00474-0.
466. Schneider, R. K. *et al.* (2014) 'Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS', *Cancer Cell*. Cell Press, 26(4), pp. 509–520. doi: 10.1016/j.ccr.2014.08.001.
467. Schnittger, S. *et al.* (2013) 'ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome', *Leukemia*. Nature Publishing Group, 27(1), pp. 82–91. doi: 10.1038/leu.2012.262.
468. Schoch, C. *et al.* (2004) 'Karyotype is an independent prognostic parameter in therapy-related acute myeloid leukemia (t-AML): An analysis of 93 patients with t-AML in comparison to 1091 patients with de novo AML', *Leukemia*. Nature Publishing Group, 18(1), pp. 120–125. doi: 10.1038/sj.leu.2403187.
469. Schoedel, K. B. *et al.* (2016) 'The bulk of the hematopoietic stem cell population is dispensable for murine steady-state and stress hematopoiesis', *Blood*. American Society of Hematology, 128(19), pp. 2285–2296. doi: 10.1182/blood-2016-03-706010.
470. Schofield, R. (1978) 'The relationship between the spleen colony-forming cell and the haemopoietic stem cell', *Blood Cells*, 4, pp. 7–25. Available at: https://pubmed.ncbi.nlm.nih.gov/747780/ (Accessed: 24 March 2021).
471. Schwaller, J. (2020) 'Learning from mouse models of MLL fusion gene-driven acute leukemia', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. Elsevier B.V., p. 194550. doi: 10.1016/j.bbagrm.2020.194550.
472. Scott, E. W. *et al.* (1994) 'Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages', *Science*. Science, 265(5178), pp. 1573–1577. doi: 10.1126/science.8079170.
473. Al Seraihi, A. F. *et al.* (2018) 'GATA2 monoallelic expression underlies reduced penetrance in inherited GATA2-mutated MDS/AML', *Leukemia*. Nature Publishing Group, 32(11), p. 1. doi: 10.1038/s41375-018-0134-9.
474. Sfeir, A. and Symington, L. S. (2015) 'Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?', *Trends in Biochemical Sciences*. Elsevier Ltd, pp. 701–714. doi: 10.1016/j.tibs.2015.08.006.

475. Shah, A. *et al.* (2013) 'Survival and cure of acute myeloid leukaemia in England, 1971-2006: a population-based study', *British Journal of Haematology*. Br J Haematol, 162(4), pp. 509–516. doi: 10.1111/bjh.12425.
476. Shalem, O. *et al.* (2014) 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science*. American Association for the Advancement of Science, 343(6166), pp. 84–87. doi: 10.1126/science.1247005.
477. Shen, M. W. *et al.* (2018) 'Predictable and precise template-free CRISPR editing of pathogenic variants', *Nature*. Nature Publishing Group, 563(7733), pp. 646–651. doi: 10.1038/s41586-018-0686-x.
478. Shen, Q. *et al.* (2018) 'Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation', *Nature*. Nature Publishing Group, 554(7690), pp. 123–127. doi: 10.1038/nature25434.
479. Shi, C. and Pamer, E. G. (2011) 'Monocyte recruitment during infection and inflammation', *Nature Reviews Immunology*, pp. 762–774. doi: 10.1038/nri3070.
480. Shimizu, R. and Yamamoto, M. (2020) 'Quantitative and qualitative impairments in GATA2 and myeloid neoplasms', *IUBMB Life*. Blackwell Publishing Ltd, 72(1), pp. 142–150. doi: 10.1002/iub.2188.
481. Shimshek, D. R. *et al.* (2002) 'Codon-improved Cre recombinase (iCre) expression in the mouse', *Genesis*, 32(1), pp. 19–26. doi: 10.1002/gene.10023.
482. Short, N. J. *et al.* (2019) 'Emerging treatment paradigms with FLT3 inhibitors in acute myeloid leukemia', *Therapeutic Advances in Hematology*. SAGE Publications, 10, p. 204062071982731. doi: 10.1177/2040620719827310.
483. Skrzeczyńska-Moncznik, J. *et al.* (2008) 'Peripheral blood CD14high CD16+ monocytes are main producers of IL-10', *Scandinavian Journal of Immunology*. John Wiley & Sons, Ltd, 67(2), pp. 152–159. doi: 10.1111/j.1365-3083.2007.02051.x.
484. Sloand, E. M. *et al.* (2007) 'CD34 cells from patients with trisomy 8 myelodysplastic syndrome (MDS) express early apoptotic markers but avoid programmed cell death by up-regulation of antiapoptotic proteins', *Blood*. The American Society of Hematology, 109(6), pp. 2399–2405. doi: 10.1182/blood-2006-01-030643.
485. Slukvin, I. I. *et al.* (2006) 'Directed Differentiation of Human Embryonic Stem Cells into Functional Dendritic Cells through the Myeloid Pathway', *The Journal of Immunology*. The American Association of Immunologists, 176(5), pp. 2924–2932. doi: 10.4049/jimmunol.176.5.2924.
486. Slukvin, I. I. (2013a) 'Deciphering the hierarchy of angiohematopoietic progenitors from human pluripotent stem cells', *Cell Cycle*. Taylor and Francis Inc., 12(5), pp. 720–727. doi: 10.4161/cc.23823.
487. Slukvin, I. I. (2013b) 'Hematopoietic specification from human pluripotent stem cells: Current advances and challenges toward de novo generation of hematopoietic stem cells', *Blood*. American Society of Hematology, pp. 4035–4046. doi: 10.1182/blood-2013-07-474825.
488. Slukvin, I. I. (2016) 'Generating human hematopoietic stem cells in vitro –exploring endothelial to hematopoietic transition as a portal for stemness acquisition', *FEBS Letters*, pp. 4126–4143. doi: 10.1002/1873-3468.12283.
489. Smith, M. A. *et al.* (2019) 'U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies', *Nature Cell Biology*. Nature Publishing Group, 21(5), pp. 640–650. doi: 10.1038/s41556-019-0314-5.
490. Snow, J. W. *et al.* (2010) 'A single cis element maintains repression of the key developmental regulator Gata2', *PLoS Genetics*. Public Library of Science, 6(9), p. e1001103. doi: 10.1371/journal.pgen.1001103.
491. Sologuren, I. *et al.* (2018) 'Lethal Influenza in Two Related Adults with Inherited GATA2 Deficiency', *Journal of Clinical Immunology*. Springer New York LLC, 38(4), pp. 513–526. doi: 10.1007/s10875-018-0512-0.
492. Song, W. J. *et al.* (1999) 'Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia', *Nature Genetics*. Nature Publishing Group, 23(2), pp. 166–175. doi: 10.1038/13793.
493. Song, X. *et al.* (2018) 'Incidence, Survival, and Risk Factors for Adults with Acute Myeloid Leukemia Not Otherwise Specified and Acute Myeloid Leukemia with Recurrent Genetic Abnormalities: Analysis of the Surveillance, Epidemiology, and End Results (SEER) Database, 2001-2013', *Acta Haematologica*. S. Karger AG, 139(2), pp. 115–127. doi: 10.1159/000486228.
494. Songdej, N. and Rao, A. K. (2017) 'Hematopoietic transcription factor mutations: Important players in inherited platelet defects', *Blood*. American Society of Hematology, pp. 2873–2881. doi: 10.1182/blood-2016-11-709881.
495. Sonoda, Y. (2008) 'Immunophenotype and functional characteristics of human primitive CD34-negative hematopoietic stem cells: The significance of the intra-bone marrow injection', *Journal of Autoimmunity*.

Academic Press, pp. 136–144. doi: 10.1016/j.jaut.2007.12.004.
496. Soukup, A. A. *et al.* (2019) 'Single-nucleotide human disease mutation inactivates a blood-regenerative GATA2 enhancer', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 129(3), pp. 1180–1192. doi: 10.1172/JCI122694.
497. Soukup, A. A. and Bresnick, E. H. (2020) 'GATA2 +9.5 enhancer: From principles of hematopoiesis to genetic diagnosis in precision medicine', *Current Opinion in Hematology*. NLM (Medline), pp. 163–171. doi: 10.1097/MOH.0000000000000576.
498. Spinner, M. A. *et al.* (2014) 'GATA2 deficiency: A protean disorder of hematopoiesis, lymphatics, and immunity', *Blood*, 123(6), pp. 809–821. doi: 10.1182/blood-2013-07-515528.
499. Spyrou, N. and Papapetrou, E. P. (2021) 'Studying leukemia stem cell properties and vulnerabilities with human iPSCs', *Stem Cell Research*. Elsevier B.V., 50. doi: 10.1016/j.scr.2020.102117.
500. Stadtfeld, M. *et al.* (2008) 'Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse', *Cell Stem Cell*. Elsevier, 2(3), pp. 230–240. doi: 10.1016/j.stem.2008.02.001.
501. Stadtfeld, M. and Hochedlinger, K. (2010) 'Induced pluripotency: History, mechanisms, and applications', *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 2239–2263. doi: 10.1101/gad.1963910.
502. Stapley, R. J. *et al.* (2021) 'Heterozygous mutation SLFN14 K208N in mice mediates species-specific differences in platelet and erythroid lineage commitment', *Blood Advances*. American Society of Hematology, 5(2), pp. 377–390. doi: 10.1182/bloodadvances.2020002404.
503. Starczynowski, D. T. *et al.* (2010) 'Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype', *Nature Medicine*. Nature Publishing Group, 16(1), pp. 49–58. doi: 10.1038/nm.2054.
504. Steensma, D. P. *et al.* (2015a) 'Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes', *Blood*. American Society of Hematology, pp. 9–16. doi: 10.1182/blood-2015-03-631747.
505. Steensma, D. P. *et al.* (2015b) 'Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes', *Blood*. American Society of Hematology, pp. 9–16. doi: 10.1182/blood-2015-03-631747.
506. Steensma, D. P. (2016) 'Graphical representation of clinical outcomes for patients with myelodysplastic syndromes', *Leukemia and Lymphoma*. Taylor and Francis Ltd, pp. 17–20. doi: 10.3109/10428194.2015.1061191.
507. Steensma, D. P. (2018a) 'Clinical consequences of clonal hematopoiesis of indeterminate potential', *Hematology (United States)*. American Society of Hematology, 2018(1), pp. 264–269. doi: 10.1182/asheducation-2018.1.264.
508. Steensma, D. P. (2018b) 'Clinical Implications of Clonal Hematopoiesis', *Mayo Clinic Proceedings*. Elsevier Ltd, pp. 1122–1130. doi: 10.1016/j.mayocp.2018.04.002.
509. Steensma, D. P. (2018c) 'Myelodysplastic syndromes current treatment algorithm 2018', *Blood Cancer Journal*. Nature Publishing Group, 8(5). doi: 10.1038/s41408-018-0085-4.
510. Steensma, D. P. (2019) 'Does early diagnosis and treatment of myelodysplastic syndromes make a difference?', *Best Practice and Research: Clinical Haematology*. Bailliere Tindall Ltd, p. 101099. doi: 10.1016/j.beha.2019.101099.
511. Stokes, N., Dominguez-Sola, D. and Papapetrou, E. P. (2018) 'Functional Analysis of MYC Deregulation By Diverse Genetic Mechanisms during Hematopoiesis', *Blood*. American Society of Hematology, 132(Supplement 1), pp. 1323–1323. doi: 10.1182/blood-2018-99-110829.
512. Stone, R. M. *et al.* (2017) 'Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 Mutation', *New England Journal of Medicine*. Massachussetts Medical Society, 377(5), pp. 454–464. doi: 10.1056/NEJMoa1614359.
513. Sturgeon, C. M. *et al.* (2014) 'Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells'. doi: 10.1038/nbt.2915.
514. Sugimoto, C. *et al.* (2015) 'Differentiation Kinetics of Blood Monocytes and Dendritic Cells in Macaques: Insights to Understanding Human Myeloid Cell Development', *The Journal of Immunology*. The American Association of Immunologists, 195(4), pp. 1774–1781. doi: 10.4049/jimmunol.1500522.
515. Sugimura, R. *et al.* (2017) 'Haematopoietic stem and progenitor cells from human pluripotent stem cells', *Nature*, 545(7655), pp. 432–438. doi: 10.1038/nature22370.
516. Sugiyama, T. *et al.* (2006) 'Maintenance of the Hematopoietic Stem Cell Pool by CXCL12-CXCR4 Chemokine Signaling in Bone Marrow Stromal Cell Niches', *Immunity*. Cell Press, 25(6), pp. 977–988. doi: 10.1016/j.immuni.2006.10.016.

517. Sumide, K. *et al.* (2018) 'A revised road map for the commitment of human cord blood CD34-negative hematopoietic stem cells', *Nature Communications*. Nature Publishing Group, 9(1), pp. 1–17. doi: 10.1038/s41467-018-04441-z.
518. Sun, J. *et al.* (2014) 'Clonal dynamics of native haematopoiesis', *Nature*. Nature Publishing Group, 514(7522), pp. 322–327. doi: 10.1038/nature13824.
519. Suzuki, M. *et al.* (2013) 'GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation', *Genes to Cells*. John Wiley & Sons, Ltd, 18(11), pp. 921–933. doi: 10.1111/gtc.12086.
520. Suzuki, M., Shimizu, R. and Yamamoto, M. (2011) 'Transcriptional regulation by GATA1 and GATA2 during erythropoiesis', *International Journal of Hematology*. Springer, 93(2), pp. 150–155. doi: 10.1007/s12185-011-0770-6.
521. Symington, L. S. (2016) 'Mechanism and regulation of DNA end resection in eukaryotes', *Critical Reviews in Biochemistry and Molecular Biology*. Taylor and Francis Ltd, pp. 195–212. doi: 10.3109/10409238.2016.1172552.
522. Taeubner, J. *et al.* (2018) 'Penetrance and Expressivity in Inherited Cancer Predisposing Syndromes', *Trends in Cancer*. Cell Press, pp. 718–728. doi: 10.1016/j.trecan.2018.09.002.
523. Taghon, T., Yui, M. A. and Rothenberg, E. V. (2007) 'Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3', *Nature Immunology*. Nat Immunol, 8(8), pp. 845–855. doi: 10.1038/ni1486.
524. Takahashi, K. *et al.* (2007) 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors', *Cell*, 131(5), pp. 861–872. doi: 10.1016/j.cell.2007.11.019.
525. Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.
526. Takai, J. *et al.* (2013) 'The Gata1 5′ region harbors distinct cis-regulatory modules that direct gene activation in erythroid cells and gene inactivation in HSCs', *Blood*. American Society of Hematology, 122(20), pp. 3450–3460. doi: 10.1182/blood-2013-01-476911.
527. Takizawa, H., Boettcher, S. and Manz, M. G. (2012) 'Demand-adapted regulation of early hematopoiesis in infection and inflammation', *Blood*. American Society of Hematology, pp. 2991–3002. doi: 10.1182/blood-2011-12-380113.
528. Talati, C., Sallman, D. and List, A. (2017) 'Lenalidomide: Myelodysplastic syndromes with del(5q) and beyond', *Seminars in Hematology*. W.B. Saunders, pp. 159–166. doi: 10.1053/j.seminhematol.2017.06.003.
529. Tanaka, T. N. and Bejar, R. (2019) 'MDS overlap disorders and diagnostic boundaries', *Blood*. American Society of Hematology, pp. 1086–1095. doi: 10.1182/blood-2018-10-844670.
530. Tang, X. *et al.* (2017) 'Runt-related transcription factor 1 regulates LPS-induced acute lung injury via NF-kB signaling', *American Journal of Respiratory Cell and Molecular Biology*. American Thoracic Society, 57(2), pp. 174–183. doi: 10.1165/rcmb.2016-0319OC.
531. Tawana, K. *et al.* (2015) 'Disease evolution and outcomes in familial AML with germline CEBPA mutations', *Blood*. American Society of Hematology, 126(10), pp. 1214–1223. doi: 10.1182/blood-2015-05-647172.
532. Tawana, K. *et al.* (2017) 'Familial CEBPA-mutated acute myeloid leukemia', *Seminars in Hematology*. W.B. Saunders, pp. 87–93. doi: 10.1053/j.seminhematol.2017.04.001.
533. Taylor, J. and Lee, S. C. (2019) 'Mutations in spliceosome genes and therapeutic opportunities in myeloid malignancies', *Genes Chromosomes and Cancer*. Blackwell Publishing Inc., pp. 889–902. doi: 10.1002/gcc.22784.
534. Tefferi, a and Vardiman, J. W. (2008) 'Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms.', *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(1), pp. 14–22. doi: 10.1038/sj.leu.2404955.
535. Teshigawara, R. *et al.* (2017a) 'Mechanism of human somatic reprogramming to iPS cell', *Laboratory Investigation*. Nature Publishing Group, pp. 1152–1157. doi: 10.1038/labinvest.2017.56.
536. Teshigawara, R. *et al.* (2017b) 'Mechanism of human somatic reprogramming to iPS cell', *Laboratory Investigation*. Nature Publishing Group, pp. 1152–1157. doi: 10.1038/labinvest.2017.56.
537. Thomson, J. A. (1998) 'Embryonic stem cell lines derived from human blastocysts', *Science*. American Association for the Advancement of Science, 282(5391), pp. 1145–1147. doi: 10.1126/science.282.5391.1145.
538. Tikhonova, A. N. *et al.* (2019) 'The bone marrow microenvironment at single-cell resolution', *Nature*. Nature Publishing Group, 569(7755), pp. 222–228. doi: 10.1038/s41586-019-1104-8.

539. Tiley, S. and Claxton, D. (2013) 'Clofarabine in the treatment of acute myeloid leukemia in older adults', *Therapeutic Advances in Hematology*. SAGE Publications, 4(1), pp. 5–13. doi: 10.1177/2040620712461666.
540. Timchenko, N. A. *et al.* (1996) 'CCAAT/enhancer-binding protein α (C/EBPα) inhibits cell proliferation through the p21 (WAF-1/CIP-1/SDI-1) protein', *Genes and Development*, 10(7), pp. 804–815. doi: 10.1101/gad.10.7.804.
541. Timmermans, F. *et al.* (2009) 'Generation of T Cells from Human Embryonic Stem Cell-Derived Hematopoietic Zones', *The Journal of Immunology*, 182(11), pp. 6879–6888. doi: 10.4049/jimmunol.0803670.
542. Ting, C. N. *et al.* (1996) 'Transcription factor GATA-3 is required for development of the T-cell lineage', *Nature*. Nature Publishing Group, 384(6608), pp. 474–475. doi: 10.1038/384474a0.
543. Tipping, A. J. *et al.* (2009) 'High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle', *Blood*, 113(12), pp. 2661–2672. doi: 10.1182/blood-2008-06-161117.The.
544. Tokoyoda, K. *et al.* (2004) 'Cellular niches controlling B lymphocyte behavior within bone marrow during development', *Immunity*. Cell Press, 20(6), pp. 707–718. doi: 10.1016/j.immuni.2004.05.001.
545. Tong, Q. *et al.* (2000) 'Function of GATA transcription factors in preadipocyte-adipocyte transition', *Science*. Science, 290(5489), pp. 134–138. doi: 10.1126/science.290.5489.134.
546. Tong, Q. *et al.* (2005) 'Interaction between GATA and the C/EBP Family of Transcription Factors Is Critical in GATA-Mediated Suppression of Adipocyte Differentiation', *Molecular and Cellular Biology*. American Society for Microbiology, 25(2), pp. 706–715. doi: 10.1128/mcb.25.2.706-715.2005.
547. Tremblay, M., Sanchez-Ferras, O. and Bouchard, M. (2018) 'Gata transcription factors in development and disease', *Development (Cambridge)*, 145(20), pp. 1–20. doi: 10.1242/dev.164384.
548. Tsai, B. F.-Y. and Orkin, S. H. (1997) 'Transcription Factor GATA-2 Is Required for Proliferation/Survival of Early Hematopoietic Cells and Mast Cell Formation, But Not for Erythroid and Myeloid Terminal Differentiation', *Blood*. American Society of Hematology, 89(10), pp. 3636–36243. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8562971 (Accessed: 13 January 2018).
549. Tsai, F.-Y. Y. *et al.* (1994) 'An early haematopoietic defect in mice lacking the transcription factor GATA-2', *Nature*. Nature Publishing Group, 371(6494), pp. 221–226. doi: 10.1038/371221a0.
550. Tulpule, A. *et al.* (2013) 'Pluripotent stem cell models of shwachman-diamond syndrome reveal a common mechanism for pancreatic and hematopoietic dysfunction', *Cell Stem Cell*. Elsevier, 12(6), pp. 727–736. doi: 10.1016/j.stem.2013.04.002.
551. Turhan, A. *et al.* (2019) 'Modeling malignancies using induced pluripotent stem cells: from chronic myeloid leukemia to hereditary cancers', *Experimental Hematology*. Elsevier Inc., 71, pp. 61–67. doi: 10.1016/j.exphem.2019.01.003.
552. Tyner, J. W. *et al.* (2018) 'Functional genomic landscape of acute myeloid leukaemia', *Nature*. Nature Publishing Group, 562(7728), pp. 526–531. doi: 10.1038/s41586-018-0623-z.
553. Uenishi, G. *et al.* (2014) 'Tenascin C promotes hematoendothelial development and T lymphoid commitment from human pluripotent stem cells in chemically defined conditions', *Stem Cell Reports*, 3(6), pp. 1073–1084. doi: 10.1016/j.stemcr.2014.09.014.
554. Vakulskas, C. A. and Behlke, M. A. (2019) 'Evaluation and reduction of crispr off-target cleavage events', *Nucleic Acid Therapeutics*. Mary Ann Liebert Inc., 29(4), pp. 167–174. doi: 10.1089/nat.2019.0790.
555. Vardiman, J. (2012) 'The classification of MDS: From FAB to WHO and beyond', *Leukemia Research*. Pergamon, pp. 1453–1458. doi: 10.1016/j.leukres.2012.08.008.
556. Velten, L. *et al.* (2017) 'Human haematopoietic stem cell lineage commitment is a continuous process', *Nature Cell Biology*. Nature Publishing Group, 19(4), pp. 271–281. doi: 10.1038/ncb3493.
557. Venezia, T. A. *et al.* (2004) 'Molecular signatures of proliferation and quiescence in hematopoietic stem cells', *PLoS Biology*. Edited by Bing Lim. Public Library of Science, 2(10), p. e301. doi: 10.1371/journal.pbio.0020301.
558. Verovskaya, E. V., Dellorusso, P. V. and Passegué, E. (2019) 'Losing Sense of Self and Surroundings: Hematopoietic Stem Cell Aging and Leukemic Transformation', *Trends in Molecular Medicine*. Elsevier Ltd, pp. 494–515. doi: 10.1016/j.molmed.2019.04.006.
559. Vinh, D. C. *et al.* (2010) 'Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia', *Blood*. American Society of Hematology, 115(8), pp. 1519–1529. doi: 10.1182/blood-2009-03-208629.
560. Virchow, R. (1989) 'As Based upon Physiological and Pathological Histology: cellular pathology.', *Nutrition Reviews*. Oxford Academic, 47(1), pp. 23–25. doi: 10.1111/j.1753-4887.1989.tb02747.x.

561. Vodyanik, M. A. *et al.* (2005) 'Human embryonic stem cell-derived CD34+ cells: Efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential', *Blood*, 105(2), pp. 617–626. doi: 10.1182/blood-2004-04-1649.
562. Vodyanik, M. A. and Slukvin, I. I. (2007) 'Hematoendothelial Differentiation of Human Embryonic Stem Cells', *Current Protocols in Cell Biology*. Wiley, 36(1), pp. 23.6.1-23.6.28. doi: 10.1002/0471143030.cb2306s36.
563. Vodyanik, M. A., Thomson, J. A. and Slukvin, I. I. (2006) 'Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures', *Blood*. American Society of Hematology, 108(6), pp. 2095–2105. doi: 10.1182/blood-2006-02-003327.
564. Vogel, C. and Marcotte, E. M. (2012) 'Insights into the regulation of protein abundance from proteomic and transcriptomic analyses', *Nature Reviews Genetics*. Nat Rev Genet, 13(4), pp. 227–232. doi: 10.1038/nrg3185.
565. Vouillot, L., Thélie, A. and Pollet, N. (2015) 'Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases', *G3: Genes, Genomes, Genetics*. Genetics Society of America, 5(3), pp. 407–415. doi: 10.1534/g3.114.015834.
566. Wang, C. Q. *et al.* (2014) 'Disruption of Runx1 and Runx3 leads to bone marrow failure and leukemia predisposition due to transcriptional and DNA repair defects', *Cell Reports*. Elsevier B.V., 8(3), pp. 767–782. doi: 10.1016/j.celrep.2014.06.046.
567. Wang, G. *et al.* (2013) 'Critical regulation of miR-200/ZEB2 pathway in Oct4/Sox2-induced mesenchymal-to-epithelial transition and induced pluripotent stem cell generation', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(8), pp. 2858–2863. doi: 10.1073/pnas.1212769110.
568. Wang, H. and Xu, X. (2017) 'Microhomology-mediated end joining: New players join the team', *Cell and Bioscience*. BioMed Central Ltd. doi: 10.1186/s13578-017-0136-8.
569. Wang, L. *et al.* (2006) 'Derivation and characterization of hematopoietic cells from human embryonic stem cells', *Methods Mol Biol*. doi: 10.1385/1-59745-046-4:179.
570. Wang, Q. *et al.* (1996) 'The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo.', *Cell*, 87(4), pp. 697–708. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8929538 (Accessed: 29 December 2018).
571. Wang, Q. F. *et al.* (2003) 'Cell cycle inhibition mediated by the outer surface of the C/EBPα basic region is required but not sufficient for granulopoiesis', *Oncogene*, 22(17), pp. 2548–2557. doi: 10.1038/sj.onc.1206360.
572. Wang, T. *et al.* (2021a) 'Sequential CRISPR gene editing in human iPSCs charts the clonal evolution of myeloid leukemia and identifies early disease targets', *Cell Stem Cell*. Cell Press. doi: 10.1016/j.stem.2021.01.011.
573. Wang, T. *et al.* (2021b) 'Sequential CRISPR gene editing in human iPSCs charts the clonal evolution of myeloid leukemia and identifies early disease targets', *Cell Stem Cell*. Cell Press. doi: 10.1016/j.stem.2021.01.011.
574. Wang, X. *et al.* (2015) 'GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies', *Haematologica*. Ferrata Storti Foundation, 100(10), pp. e398–e401. doi: 10.3324/haematol.2015.127092.
575. Wehr, C. *et al.* (2018) *A novel disease-causing synonymous exonic mutation in GATA2 affecting RNA splicing*, *Blood*. American Society of Hematology. doi: 10.1182/blood-2018-03-837336.
576. Wei, A. *et al.* (2016) 'Safety and Efficacy of Venetoclax Plus Low-Dose Cytarabine in Treatment-Naive Patients Aged ≥65 Years with Acute Myeloid Leukemia', *Blood*. American Society of Hematology, 128(22), pp. 102–102. doi: 10.1182/blood.v128.22.102.102.
577. Wei, S. *et al.* (2018) 'Heterozygous IDH1R132H/WTcreated by "single base editing" inhibits human astroglial cell growth by downregulating YAP', *Oncogene*, pp. 1–15. doi: 10.1038/s41388-018-0334-9.
578. Weng, A. P. *et al.* (2004) 'Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia', *Science*, 306(5694), pp. 269–271. doi: 10.1126/science.1102160.
579. Wesely, J. *et al.* (2020) 'Acute Myeloid Leukemia iPSCs Reveal a Role for RUNX1 in the Maintenance of Human Leukemia Stem Cells', *Cell Reports*. Elsevier B.V., 31(9), p. 107688. doi: 10.1016/j.celrep.2020.107688.
580. West, R. R. *et al.* (2014) 'Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation', *Haematologica*. Ferrata Storti Foundation, 99(2), pp. 276–281. doi: 10.3324/haematol.2013.090217.
581. Wilson, A. *et al.* (2008) 'Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair', *Cell*. Cell Press, 135(6), pp. 1118–1129. doi:

10.1016/j.cell.2008.10.048.
582. Wisniewski, D. *et al.* (2011) 'Further phenotypic characterization of the primitive lineage− CD34+CD38−CD90+CD45RA− hematopoietic stem cell/progenitor cell sub-population isolated from cord blood, mobilized peripheral blood and patients with chronic myelogenous leukemia', *Blood Cancer Journal*. Nature Publishing Group, 1(9), p. e36. doi: 10.1038/bcj.2011.35.
583. Wlodarski, M. W. *et al.* (2016) 'Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents', *Blood*, 127(11), pp. 1387–1397. doi: 10.1182/blood-2015-09-669937.
584. Wlodarski, M. W., Collin, M. and Horwitz, M. S. (2017) 'GATA2 deficiency and related myeloid neoplasms', *Seminars in Hematology*. W.B. Saunders, 54(2), pp. 81–86. doi: 10.1053/j.seminhematol.2017.05.002.
585. Wong, K. L. *et al.* (2011a) 'Gene expression profiling reveals the defining features of the classical, intermediate,and nonclassical human monocyte subsets', *Blood*. Blood, 118(5). doi: 10.1182/blood-2010-12-326355.
586. Wong, K. L. *et al.* (2011b) 'Gene expression profiling reveals the defining features of the classical, intermediate,and nonclassical human monocyte subsets', *Blood*. American Society of Hematology, 118(5), pp. e16–e31. doi: 10.1182/blood-2010-12-326355.
587. Wu, Z. *et al.* (2019) 'Single-Cell RNA Sequencing Reveals a Distinct Transcriptome Signature of Hematopoiesis in GATA2 Deficiency', *Blood*. American Society of Hematology, 134(Supplement_1), pp. 3735–3735. doi: 10.1182/blood-2019-124774.
588. Wu, Z. *et al.* (2020) 'Sequencing of RNA in single cells reveals a distinct transcriptome signature of hematopoiesis in GATA2 deficiency', *Blood Advances*. American Society of Hematology, 4(12), pp. 2702–2716. doi: 10.1182/bloodadvances.2019001352.
589. Xu, C. *et al.* (2018) 'Stem cell factor is selectively secreted by arterial endothelial cells in bone marrow', *Nature Communications*. Nature Publishing Group, 9(1). doi: 10.1038/s41467-018-04726-3.
590. Xu, J. *et al.* (2007) 'Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(30), pp. 12377–12382. doi: 10.1073/pnas.0704579104.
591. Xu, Y. *et al.* (2009) 'Downregulation of GATA-2 and overexpression of adipogenic gene-PPARγ in mesenchymal stem cells from patients with aplastic anemia', *Experimental Hematology*. Exp Hematol, 37(12), pp. 1393–1399. doi: 10.1016/j.exphem.2009.09.005.
592. Yamamoto, R. *et al.* (2013) 'Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells', *Cell*. Cell Press, 154(5), pp. 1112–1126. doi: 10.1016/j.cell.2013.08.007.
593. Yang, J. *et al.* (2014) 'Monocyte and macrophage differentiation: Circulation inflammatory monocyte as biomarker for inflammatory diseases', *Biomarker Research*, 2(1), pp. 1–9. doi: 10.1186/2050-7771-2-1.
594. Yao, L. *et al.* (2005) 'Bone marrow dysfunction in mice lacking the cytokine receptor gp130 in endothelial cells', *Blood*. The American Society of Hematology, 106(13), pp. 4093–4101. doi: 10.1182/blood-2005-02-0671.
595. Yao, S. *et al.* (2004) 'Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications', *Molecular Therapy*. Academic Press Inc., 10(1), pp. 27–36. doi: 10.1016/j.ymthe.2004.04.007.
596. Yokota, T. *et al.* (2000) 'Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages', *Blood*. American Society of Hematology, 96(5), pp. 1723–1732. doi: 10.1182/blood.v96.5.1723.h8001723_1723_1732.
597. Yoshida, K. *et al.* (2011) 'Frequent pathway mutations of splicing machinery in myelodysplasia', *Nature*. Nature Publishing Group, 478(7367), pp. 64–69. doi: 10.1038/nature10496.
598. Yu, C. *et al.* (2002) 'Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo', *Journal of Experimental Medicine*. The Rockefeller University Press, 195(11), pp. 1387–1395. doi: 10.1084/jem.20020656.
599. Yu, P. *et al.* (2011) 'FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation', *Cell Stem Cell*. Cell Stem Cell, 8(3), pp. 326–334. doi: 10.1016/j.stem.2011.01.001.
600. Yu, V. W. C. *et al.* (2015) 'Specific bone cells produce DLL4 to generate thymus-seeding progenitors from bone marrow', *Journal of Experimental Medicine*. Rockefeller University Press, 212(5), pp. 759–774. doi: 10.1084/jem.20141843.

601. Yu, V. W. C. *et al.* (2016) 'Distinctive Mesenchymal-Parenchymal Cell Pairings Govern B Cell Differentiation in the Bone Marrow', *Stem Cell Reports*. Cell Press, 7(2), pp. 220–235. doi: 10.1016/j.stemcr.2016.06.009.
602. Yu, V. W. C. and Scadden, D. T. (2016) 'Hematopoietic Stem Cell and Its Bone Marrow Niche', in *Current Topics in Developmental Biology*. Academic Press Inc., pp. 21–44. doi: 10.1016/bs.ctdb.2016.01.009.
603. Yue, R. *et al.* (2016) 'Leptin Receptor Promotes Adipogenesis and Reduces Osteogenesis by Regulating Mesenchymal Stromal Cells in Adult Bone Marrow', *Cell Stem Cell*. Cell Press, 18(6), pp. 782–796. doi: 10.1016/j.stem.2016.02.015.
604. Yung, S. K. *et al.* (2013) 'Brief report: Human pluripotent stem cell models of fanconi anemia deficiency reveal an important role for fanconi anemia proteins in cellular reprogramming and survival of hematopoietic progenitors', *Stem Cells*. John Wiley & Sons, Ltd, 31(5), pp. 1022–1029. doi: 10.1002/stem.1308.
605. Zaidan, N. and Ottersbach, K. (2018a) 'The multi-faceted role of Gata3 in developmental haematopoiesis', *Open Biology*. Royal Society Publishing, p. 180152. doi: 10.1098/rsob.180152.
606. Zaidan, N. and Ottersbach, K. (2018b) 'The multi-faceted role of Gata3 in developmental haematopoiesis', *Open Biology*. Royal Society Publishing. doi: 10.1098/rsob.180152.
607. Zamani, F. *et al.* (2013) 'Induction of CD14 expression and differentiation to monocytes or mature macrophages in promyelocytic cell lines: New approach', *Advanced Pharmaceutical Bulletin*. Tabriz University of Medical Sciences, 3(2), pp. 329–332. doi: 10.5681/apb.2013.053.
608. Zaret, K. S. and Carroll, J. S. (2011) 'Pioneer transcription factors: Establishing competence for gene expression', *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 2227–2241. doi: 10.1101/gad.176826.111.
609. Zawada, A. M. *et al.* (2011) 'SuperSAGE evidence for CD14 ++CD16 + monocytes as a third monocyte subset', *Blood*. Blood, 118(12). doi: 10.1182/blood-2011-01-326827.
610. Zhang, J. *et al.* (2012) 'The genetic basis of early T-cell precursor acute lymphoblastic leukaemia', *Nature*. Nature Publishing Group, 481(7380), pp. 157–163. doi: 10.1038/nature10725.
611. Zhang, L. *et al.* (2020) 'Molecular and cellular mechanisms of aging in hematopoietic stem cells and their niches', *Journal of Hematology and Oncology*. BioMed Central Ltd, pp. 1–22. doi: 10.1186/s13045-020-00994-z.
612. Zhang, M. Y. *et al.* (2015) 'Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy', *Nature Genetics*. Nature Publishing Group, 47(2), pp. 180–185. doi: 10.1038/ng.3177.
613. Zhang, Q. *et al.* (2015) 'Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6', *Nature*. Nature Publishing Group, 525(7569), pp. 389–393. doi: 10.1038/nature15252.
614. Zhang, S.-J. *et al.* (2008) 'Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(6), pp. 2076–81. doi: 10.1073/pnas.0711824105.
615. Zhang, S.-J. J. *et al.* (2009) 'GATA-2 L359 V mutation is exclusively associated with CML progression but not other hematological malignancies and GATA-2 P250A is a novel single nucleotide polymorphism', *Leukemia Research*. Elsevier, 33(8), pp. 1141–3. doi: 10.1016/j.leukres.2009.02.025.
616. Zhang, Y. *et al.* (2018) 'Hematopoietic Hierarchy – An Updated Roadmap', *Trends in Cell Biology*. Elsevier Ltd, pp. 976–986. doi: 10.1016/j.tcb.2018.06.001.
617. Zhou, B. O. *et al.* (2017) 'Bone marrow adipocytes promote the regeneration of stem cells and haematopoiesis by secreting SCF', *Nature Cell Biology*. Nature Publishing Group, 19(8), pp. 891–903. doi: 10.1038/ncb3570.
618. Zhou, Y. *et al.* (2019) 'Overexpression of GATA2 Enhances Development and Maintenance of Human Embryonic Stem Cell-Derived Hematopoietic Stem Cell-like Progenitors', *Stem Cell Reports*. Cell Press, 13(1), pp. 31–47. doi: 10.1016/j.stemcr.2019.05.007.
619. Zhu, J. *et al.* (2007) 'Osteoblasts support B-lymphocyte commitment and differentiation from hematopoietic stem cells', *Blood*. Blood, 109(9), pp. 3706–3712. doi: 10.1182/blood-2006-08-041384.