Genetics and Functional Characterization of GATA2, a Novel Cancer Gene in Familial Leukaemia

CHAN ENG CHONG

Bachelor of Science (Hons.) (Biochemistry)

Master of Science (Biochemistry)

A thesis submitted for the Degree of Doctor of Philosophy

School of Medicine

(Discipline of Medicine)

Faculty of Health Sciences

University of Adelaide, South Australia.

July 2013

TABLE OF CONTENTS

ABSTRACT vii

STATEMENT ix

ACKNOWLEDGMENTS xi

LIST OF FIGURES xiii

LIST OF TABLES xv

LIST OF ABBREVIATIONS xvi

Chapter 1: Literature Review

1.1 Introduction to Haematopoiesis 1

1.1.1 Ontology of the Haematopoietic System 2

1.1.2 Transcription Factors in Myelopoiesis 5
1.1.2.1 Haematopoietic Stem Cell Transcription Factors 6
1.1.2.2 Common Myeloid Progenitors and Granulocyte Macrophage Progenitors 9
1.1.2.3 Megakaryocyte Erythroid Progenitors 9

1.2 Leukaemias and Other Myeloid Malignancies 10

1.2.1 Classification of Acute Myeloid Leukaemia 10
1.2.1.1 French-American-British System (FAB) 11
1.2.1.2 WHO Classification 12

1.2.2 The Genetics of Myeloid Malignancies 13
1.2.2.1 Intermediate and Small Genetic Lesions 14
1.2.2.2 Large Genetic Lesions 15

1.2.3 Acute Myeloid Leukaemia and Myeloid Neoplasms 16
1.2.3.1 *De novo* AML 17
1.2.3.2 Myelodysplastics Syndromes 18
1.2.3.3 Myeloproliferative Neoplasms 19

1.2.4 Pure Familial Leukaemia 21
1.2.4.1 *RUNX1* Associated Familial Platelet Disorder with Propensity to AML 21

1.2.4.2 *CEBPA* Associated Leukaemia 21
1.2.4.3 *GATA2* Associated Myeloid Malignancies 22

1.2.5 Hereditary Syndromes with Predisposition to Leukaemia 22
1.2.5.1 Emberger Syndrome 24
1.2.5.2 MonoMAC/DCML Deficiency Syndrome 24

1.3 Introduction to GATA Family of Transcription Factors 25

1.3.1 GATA Family Members 26
1.3.1.1 GATA1 26
1.3.1.2 GATA2 28
1.3.1.3 GATA3 31

1.3.2 GATA1 and 2 Related Diseases 31

1.4 Aim of the Study 33

1.5 References 34

Chapter 2: Materials and Methods

2.1 Introduction 49

2.2 General Methods 50
2.2.1 DNA Isolation 50
2.2.2 Total RNA Isolation 50
2.2.3 DNA or RNA Purification 50
2.2.4 Restriction Enzyme Digestion 50
2.2.5 DNA Dephosphorylation and Cloning 51
2.2.6 Routine PCR 51
2.2.7 Colony PCR 52
2.2.8 Site-Directed Mutagenesis 52
2.2.9 Tissue Culture 53
2.2.10 Lipofectamine-Mediated Transfection 53
2.2.11 Whole Cell Lysate Preparation 53
2.2.12 Nuclear Lysate Preparation 54

2.3 Chapter 3 Methods 54

2.4 Chapter 4 Methods 54
2.4.1 Generation of *GATA2* Mutant Constructs 54
2.4.2 Luciferase Reporter Assay 54
2.4.3 Homology Modelling of GATA2 WT and Mutants 55
2.4.4 Verification of GATA2 Theoretical Model 55

2.4.5 Generation of FLAG-*GATA2* Expression Constructs 55
2.4.6 Generation of Inducible FLAG-*GATA2* Dual Lentiviral Expression Constructs 55
2.4.7 Generation of Regulatable FLAG-*GATA2* Expressing HL-60 Cell Lines 56
2.4.8 Co-Immunoprecipitation 56
2.4.9 Generation of *Gata2* Retrovirus Constructs 56
2.4.10 Generation of Retrovirus 57
2.4.11 Animal Handling and Bone Marrow Cell Extraction 57
2.4.12 Retrovirus Transduction and Colony forming Unit Assay 57

2.5 Chapter 5 Methods 58
2.5.1 Generation of *Gata2* Mutant Constructs 58
2.5.2 Immunofluorescence Staining 58
2.5.3 Generation of *PEE* GATA Mutated Site Luciferase Reporter Plasmid 58

2.6 Media and Solutions 58

2.7 References 61

Chapter 3: Novel GATA2 Mutations in Familial MDS/AML

3.1 Summary 62

3.2 Notes 63

3.3 Permission to Reuse Published Materials 64

3.4 The Published Article 64

3.5 Additional Results 73

3.6 General Discussion 76
3.6.1 GATA2 is a New Predisposition Gene for MDS and AML 76
3.6.2 Role of GATA2 in Haematological Disease 77

3.7 Conclusion 79

3.8 References 79

Chapter 4: Characterization of Novel GATA2 Mutations Associated with Haematological Malignancies, Immunodeficiency Syndrome and Lymphoedema

4.1 Introduction 81

4.2 Results 83

4.2.1 GATA2 Mutants Appropriately Localize to the Nucleus 83
4.2.2 GATA2 Mutants Reduce Transactivation 84
4.2.3 GATA2 Mutants Reduce DNA Binding Potential 86
4.2.4 Homology and Structural Modelling of GATA2 and Mutants 88
4.2.5 GATA2 Mutants Alter Protein-Protein Interaction 91
4.2.6 GATA2 Mutants Allow Myeloid Progenitor Differentiation *Ex Vivo* 93
4.2.7 GATA2 Mutants Do Not Confer Self-Renewal Capacity *Ex Vivo* 95

4.3 Discussion 98
4.3.1 Substitution of Conserved Residues in GATA2 Disrupt Its Function 98
4.3.2 GATA2 Mutants Cause Aberrant Protein Partnerships 100
4.3.3 GATA2 Mutants Display Loss-of-Function in Haematopoietic Cells 100
4.3.4 T354M is a Potent Mutation Affecting Myeloid Differentiation 102

4.4 Conclusion 105

4.5 Recommendations for Future Work 105

4.6 References 107

Chapter 5: Effect of GATA2 Mutations in the Lymphatic System

5.1 Introduction 109

5.2 Results 111
5.2.1 GATA2 Transactivates the *Prox1* Promoter/Enhancer Element 111
5.2.2 GATA2 Binds to *Prox1* Promoter/Enhancer Element 113
5.2.3 GATA2 WT and Mutants Exhibit Differential Binding Affinity to GATA2-Responsive Elements in *PEE* and Human Haematopoietic Lineage Promoters 115
5.2.4 *Prox1* -11.3kb Enhancer is Transactivated by Multiple Transcription Factors 116

5.3 Discussion 119
5.3.1 GATA2 is regulator for *Prox1* expression 119
5.3.2 *Prox1* Enhancer is a Functioning *cis*-Regulatory Element 120

5.4 Conclusion 121

5.5 Recommendations for Future Work 122

5.5 References 122

Chapter 6: Final Discussion and Conclusion

6.1 GATA2 is a Predisposition Gene for MDS/AML, Infectious Diseases and Emberger Syndrome 125

6.2 GATA2 in Sporadic AML and CML 130

6.3 Significance and Impact of the GATA2 Discovery 134

6.4 Conclusion 135

6.5 References 135

Appendix A Additional Information for Chapter 3 138

A-1 STATEMENT OF AUTHORSHIP 138

A-2 SUPPLEMENTARY METHODS 145

A-3 SUPPLEMENTARY INFORMATION 1 150

A-4 SUPPLEMENTARY NOTE 169

A-5 SUPPLEMENTARY INFORMATION 2 173

Appendix B Additional Information for Chapter 4 174

B-1 STATEMENT/DECLARATION OF CONTRIBUTION 174

B-2 SUPPLEMENTARY INFORMATION 176

Appendix C Additional Information for Chapter 5 178

C-1 STATEMENT/DECLARATION OF CONTRIBUTION 178

C-2 SUPPLEMENTARY INFORMATION 1 179

C-3 SUPPLEMENTARY INFORMATION 2 181

Supplementary References 182

ABSTRACT

We first report *GATA2* mutations (heterozygous) in 4 families that are susceptible to MDS/AML (3 large families) and MDS (1 small family). Molecular analysis revealed a germline transmission of a *GATA2* missense mutation (T354M) in MDS/AML families and a *GATA2* deletion mutation (T355del) in MDS family. Neither germline *RUNX1* nor *CEBPA* mutations were found in these families, in 695 non-leukemic ethnically matched controls and 268 sporadic AML samples. The mutations resided within the GATA2 zinc finger 2 domain, a critical region for DNA-binding and protein-protein interactions, but not for nuclear localization. T354M reduced DNA binding ability of GATA2; whereas, T355del bound very little, if any, to the consensus WGATAR DNA motif. T354M and T355del also significantly reduced the transactivation of GATA2 in known GATA2 responsive sequences. Moreover, co-transfection of T354M or T355del with WT reduced WT transactivation ability, suggesting that these mutants act in a dominant negative fashion. Regulatable stable promyelocytic HL-60 cells expressing WT and mutants were generated. Forced expression of WT and T354M inhibited HL-60 cell differentiation when induced with all *trans* retinoic acid. However, when compared to WT, T354M enabled cell proliferation/survival while simultaneously reducing apoptosis. In contrast, T355del was a complete loss-of-function mutant. Microarray studies elucidated that both T354M and T355del significantly decreased the expression of downstream target genes. Together, our data suggest that both T354M and T355del are loss-of-function mutations with some dominant negative attributes.

Recently, we and others have described *GATA2* genetic lesions in other diseases. We further investigated *in vitro* functions of an alllelic series of GATA2 mutants representing the major disease phenotypes: MDS/AML (T354M), MDS (T355del), CML-BC (L359V), Emberger syndrome (R361L and C373R), AML-M5 and biallellic *CEBPA* AML (R362Q), and immunodeficiency syndrome (R398W). We showed that these GATA2 mutants (except L359V) are loss-of-function that reduce DNA binding affinity and transactivation of target genes. Nevertheless, they maintain the ability to bind to known protein binding partners. Intriguingly, T354M and C373R have an enhanced affinity for PU.1, highlighting that these

mutants can influence both DNA-binding and protein-protein interaction. Preliminary transduction of *Gata2* WT or mutant expression constructs into mouse whole bone marrow cells demonstrated that GATA2 mutants did not confer self-renewal capacity, but allowed specific myeloid progenitor differentiation.

We further demonstrated that *Gata2* is expressed in lymphatic endothelial cells and that it can bind to and transactivate a *Prox1* promoter/enhancer element (*PEE*) region. *Prox1* is required for lymphatic development and maintenance, and hence *Gata2* may contribute to lymphoedema throught its action on *Prox1*. Intriguingly, *Gata2* mutants displayed differential binding affinity to two GATA binding sites and reduced transactivation of the *PEE* region. Furthermore, an enhancer region 11.3kb upstream of *Prox1* is activated by GATA2, FOXC2 and SOX18, but repressed by PROX1 itself suggesting that these key lymphatic TFs may co-operate to regulate *Prox1* expression.

In conclusion, I present the experimental work for the landmark discovery of a new MDS/AML predisposition gene. I have also characterized the molecular landscape of *GATA2* mutations where each of the mutations confers specific and major effects on GATA2 function, but where there are also subtle differences between the mutants in the contexts of DNA binding and transactivation.

STATEMENT

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. The author acknowledges that copyright of published works (as listed in next page*) contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed,

CHAN ENG CHONG

Date: 29 JULY 2013

* THE PUBLISHED ARTICLES

Methods published in the following article were cited in Chapter 3.

Hahn CN, **Chong CE**, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, Babic M, Lin M, Carmagnac A, Lee YK, Kok CH, Gagliardi L, Friend KL, Ekert PG, Butcher CM, Brown AL, Lewis ID, To LB, Timms AE, Storek J, Moore S, Altree M, Escher R, Bardy PG, Suthers GK, D'Andrea RJ, Horwitz MS, Scott HS. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* 2011 Sep 4;43(10):1012-7.

Materials, methods and partial results published in the following article were cited in Chapter 4.

Kazenwadel J, Secker GA, Liu YJ, Rosenfeld JA, Wildin RS, Cuellar-Rodriguez J, Hsu AP, Dyack S, Fernandez CV, **Chong CE**, Babic M, Bardy PG, Shimamura A, Zhang MY, Walsh T, Holland SM, Hickstein DD, Horwitz MS, Hahn CN, Scott HS, Harvey NL. *Blood*. 2012 Feb 2;119(5):1283-91.

ACKNOWLEDGMENTS

The work presented in this thesis was performed in the Molecular Pathology Research Laboratory at the in the Department of Molecular Pathology, in the Centre for Cancer Biology (a centre of excellence in cancer research) of SA Pathology under the supervision of Prof. Hamish S. Scott and Dr Christopher N. Hahn.

First of all, I would like to thank my supervisor, Prof. Scott for giving me an opportunity and accepting me as a PhD. student in his laboratory. Throughout my PhD, I have been granted the freedom to devise, conduct and implement my experiments. Through this freedom, he moulded me from a naïve student to an independent researcher. He reminds me of Albert Einstein who famously said, "I never teach my pupils; I only attempt to provide the conditions in which they can learn." In other words, he constantly inspires me to think out of the box and see things from multiple angles. Hamish also challenges and motivates me to step out of my comfort zone and encourages me to explore the new realm for the answers I am seeking. Hence, my growth as a scientist in both scientific knowledge and confidence are well looked after in his laboratory. I would also like to take this opportunity to thank him for his generous financial aid in supporting my travel for attending both national and international conferences, hence improving my scientific knowledge and allowing me to interact with many brilliant minds. Thank you Hamish!

I would like to express my gratitude to my co-supervisor, Dr Hahn, who has been advising, motivating and guiding me through my PhD. Chris is a knowledgeable (a.k.a. walking dictionary) and a patient man. He taught me basically everything that I needed to know for my study. I have enjoyed very much the theoretical framework discussion in his new office and I learned a lot from him. Chris also taught me to interpret, analyse, clarify and articulate my available data into the strongest and persuasive arguments. Here, I would also like to thank him for his patience and time he spent on reading and correcting my thesis written in Manglish (Malaysian English) and his efforts in improving my writing skills. Thank you Chris!

I would like to express my thanks for the great time and support to all who work in the Scott laboratory: Mr Peter Brautigan, Ms Milena Babic, Dr Manuela Klingler-Hoffmann and Ms Parvathy Venugopal. Special thanks to Ms Young Kyun Lee for her dauntless and persistent efforts in the proteomic work and Dr David Yeung (haematologist) for his professional opinion and constructive suggestion for my thesis. Without their help, my PhD. journey would not have been smooth. I would also like to thank all my ex-lab members for the quality time we spent during the course of my study, particularly Dr Michael Ling King Hwa, Dr Brita Ardesjo Lungren, Mr Scott Warming and Mr Joseph Carolan.

Special thanks particularly go to all our collaborators: A/Prof. Richard D'Andrea, Dr Anna Brown, Dr Chung Kok, Dr Sonya Diakiw, Mr Grant Engler and Ms Diana Larossi (Acute Leukaemia Laboratory) and Dr Genevieve Secker, Dr Drew Sutton, Ms Jan Kazenwadel and Dr Natasha Harvey (Lymphatic Development Laboratory). Without their help, I could not have survived this PhD. I would like to thank all Molecular Pathology staff, particularly Dr John Rayner and Ms Ming Lin for their technical contribution toward the completion of my experiments and data analysis. Several groups have also generously allowed me to use their equipment over the years, including A/Prof. Richard D'Andrea (cytospin), Prof. Gregory Goodall and Dr Yeesim Khew-Goodall (confocal microscope and multiwell-plate luminometers) and Dr Jane Arthur (Ultracentrifuge).

Above all, I must thank my family. No other words could express how truly grateful and indebted I am to my parents and siblings. I thank them for their endless patience and for their support during my study. On a personal note, I would like to thank Ms Sze Yee Phuah. I would not be here without her love, patience, encouragement and support. She shares my greatest disappointments along with my successes. Thank you very much Sze Yee!

LIST OF FIGURES

Figure 1.1:	A Classic Model of the Hierarchical Development of Haematopoietic Cells.	2
Figure 1.2:	Conventional and Alternative Models for Haematopoietic Stem Cell and Blood Lineage Commitment in Murine.	4
Figure 1.3:	A Revised Version of the Classical Model of the Haematopoietic Tree.	5
Figure 1.4:	Model of GATA2 Function in the Haematopoietic Lineage Determination.	8
Figure 1.5:	Schematic of Heptad Proteins Bound to DNA.	8
Figure 1.6:	AML-Related Genetic Pathways.	17
Figure 1.7:	The Clonal Bone Marrow Disorders.	20
Figure 3.1:	Binding of GATA2 ZF2 to the *GM-CSF* and *Prox1* Oligonucleotide Probes Analyzed by EMSA.	74
Figure 4.1:	Expression of GATA2 Mutants in HEK293 Cells.	83
Figure 4.2:	Transactivation of GATA2 on Known Responsive Elements.	85
Figure 4.3:	DNA Binding Assay on GATA2 WT and Mutant Proteins.	87
Figure 4.4:	Multiple Sequence Alignment of GATA2 ZF2 Motif Across Different Vertebrates and Among GATA Family Members.	89
Figure 4.5:	Structural Modelling of GATA2 WT and Mutants.	90
Figure 4.6:	Co-Immunoprecipitation of GATA Proteins and PU.1.	92
Figure 4.7:	Whole BM Clonogenic Assays.	94
Figure 4.8:	Clonogenic Assays on Total Bone Marrow Cells Transduced with Gata2 WT or Mutants.	96
Figure 4.9:	Participation of GATA2 in the Haematopoietic Transcriptional Regulatory Network.	103
Figure 4.10:	Gain-of-Function and Loss-of-Function Models for GATA2 Mutation.	104
Figure 5.1:	*Prox1* Enhancer/Promoter Elements Contain Functional GATA Sites.	112

Figure 5.2: Functional Characterization of *PEE*. 114

Figure 5.3: WEMSA for GATA2 WT and Mutants on a Panel of Known Responsive Promoter/Enhancer Elements. 116

Figure 5.4: WEMSA and Luciferase Reporter Assays on the *Prox1* -11.3kb Enhancer. 118

Figure 6.1: Distribution of *GATA2* Mutations Identified in Various Haematological Malignancies, MonoMAC/DCML Deficiency and Emberger Syndrome. 129

Figure 6.2: GATA2 Mutant Proteins According to Germline or Somatic Mutation. 133

LIST OF TABLES

Table 1.1:	Frequency of FAB Subtypes in a Study Cohort Consists of 614 Patients with *de novo* AML.	11
Table 1.2:	A Revised WHO Classification of AML.	13
Table 1.3:	Genetic Factors Predisposing to the Development of Secondary AML.	23
Table 2.1:	List of Restriction Enzymes.	51
Table 3.1:	Association Constants (Ka) and Dissociation Constants (Kd) Determined by Isothermal Titration Calorimetry (ITC) for Each GATA2 Constructs Using the *GM-CSF* Probe.	75
Table 4.1:	Colony-Forming Activities of GFP-Sorted BM Cells.	94
Table 4.2:	Summary of the *In Vitro* Characterization of GATA2 Mutants.	97

LIST OF ABBREVIATIONS

Abbreviations	Description
4HT	4-hydroxytamoxifen
ALL	acute lymphoid leukaemia
ATRA	all *trans* retinoic acid
BFU-E	burst forming unit-erythroid
BM	bone marrow
CFU	colony forming unit
CFU-G	CFU-Granulocyte
CFU-GEMM	CFU-Granulocyte/Erythrocyte/Monocyte/Megakaryocyte
CFU-GM	CFU-Granulocyte/Macrophage
CFU-M	CFU-Macrophage
ChIP	chromatin immunoprecipitation
ChIP-Seq	chromatin immunoprecipitation-sequencing
CLL	chronic lymphoid leukaemia
CLP	common lymphoid progenitor
CML-BC	chronic myeloid leukaemia blast crisis
CMP	common myeloid progenitor
CN-AML	cytogenetically normal AML
Co-IP	co-immunoprecipitation
EMSA	electrophoretic mobility shift assay
ENCODE	ENCyclopedia of DNA Elements
ES cells	Embryonic stem cells
FPD/AML	familial thrombocytopenia with increased risk to develop AML
G-CSF	granulocyte-colony stimulating factor
GFP	green fluorescence protein
GM-CSF	granulocyte-macrophage colony-stimulating factor
GMP	granulocyte-macrophage progenitor

GOF	gain-of-function
HM	haematological malignancies
HPCs	haematopoietic progenitor cells
HSCs	haematopoietic stem cells
KO	knock-out
LEC	lymphatic endothelial cells
LOF	loss-of-function
LSK cells	Lineage negative, SCAL positive, c-KIT positive cells
MDS	myelodysplastic syndrome
MDS/AML	AML with myelodysplasia-related changes
MPN	myeloproliferative neoplasm
OMIM	Online Mendelian Inheritance in Man
PB	peripheral blood
RA	refractory anaemia
REAB-1	refractory anaemia with excess blasts 1
REAB-2	refractory anaemia with excess blasts 2
SEM	standard error mean
SCF	stem cell factor
t-AML	therapy-related AML
TF	transcription factor
TPO	thrombopoietin
TCRD	T cell Receptor Delta
UCSC	University of California, Santa Cruz
WEMSA	Western blotting-electrophoretic mobility shift assay
WT	wild type
ZF1	zinc finger 1
ZF2	zinc finger 2

Chapter 1: Literature Review

1.1 Introduction to Haematopoiesis

Haematopoiesis is the formation of many types of blood cells with specialized functions in the bone marrow. Blood formation is an ongoing process throughout life. Each day billion of new haematopoietic cells are generated to replenish aged blood cells. The sustainability of new blood cell production essentially relies on the presence of a primary source know as haematopoietic stem cells (HSCs). The major function of the HSC is to maintain blood homeostasis by proliferating and differentiating pluripotent stem cells into non-dividing and short-lived mature cells, a process known as lineage commitment. Broadly, HSC can differentiate into lymphoid and myeloid lineages (**Figure 1.1**). The myeloid lineage consists of erythrocytes, platelets, monocytes, macrophages and granulocytes. Whereas, lymphoid lineage is mostly comprised of immune cells such as T and B cells, natural killer cells and dendritic cells.

Haematopoietic development is a complex process that is tightly regulated by a number of intrinsic (transcription factors) and extrinsic (cytokines and growth factors) factors [1]. Deregulation of haematopoiesis caused by genetic defects or genotoxic agents may lead to diseases, ranging from mild form of anaemia, neutropaenia and thrombocytopaenia, to more severe and life-threatening illness such as multiple myeloma, lymphomas and myelodysplastic syndrome/acute myeloid leukaemia (MDS/AML). Understanding the aetiology and genetic mechanisms underlying abnormal haematopoiesis will help to tackle these diseases and possibly identify novel therapeutic targets.

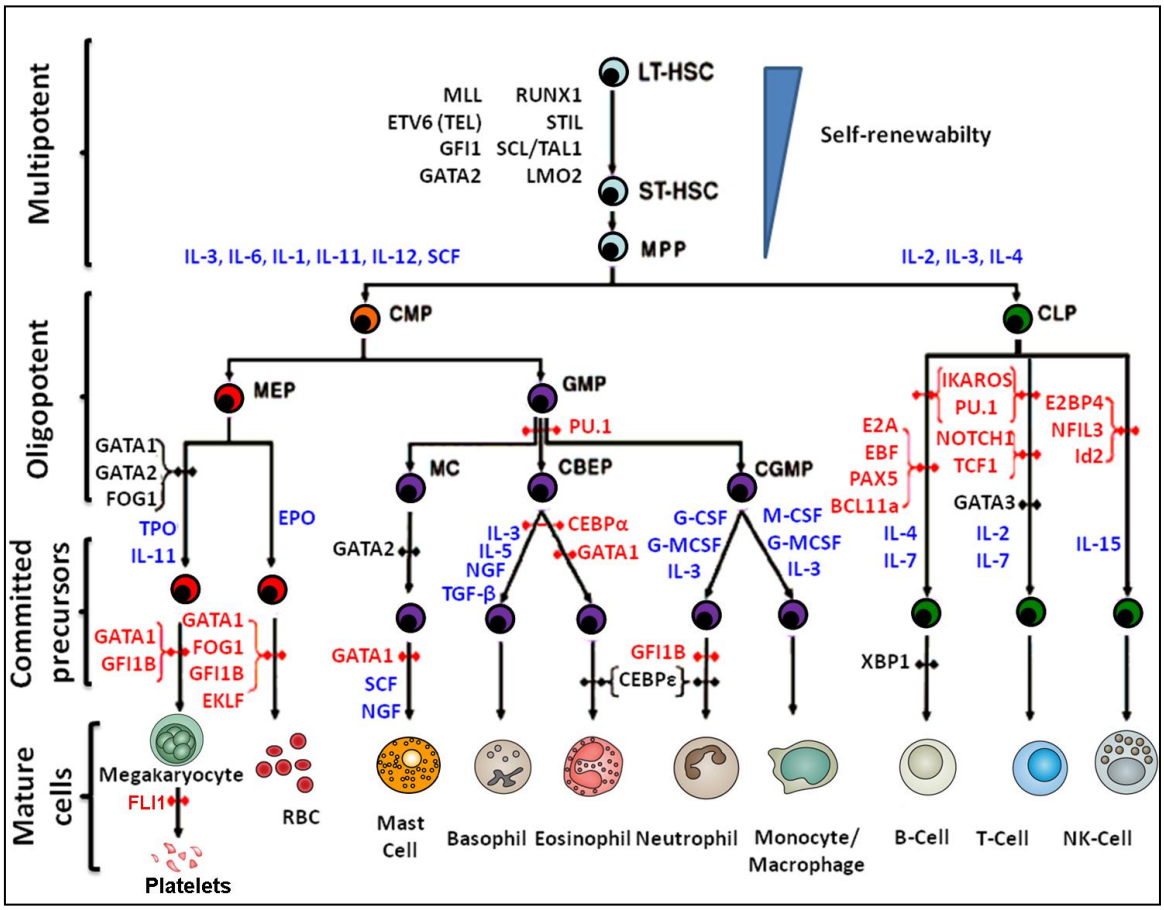


Figure 1.1: Classic Model of the Hierarchical Development of Haematopoietic Cells.

A schematic diagram depicting key haematopoietic transcription factors (TFs) and cytokines that is required for HSC formation, survival, differentiation and function. TF in red indicate the stages at which haematopoietic development is blocked in the absence of a given TF, as determined through conventional gene knockouts. TF in black has been associated with oncogenesis. Key cytokines required for proper development are indicated in blue. Abbreviations: LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/macrophage progenitor; MC, mast cell; CEBP, common eosinophil/basophil progenitor; CGMP, common granulocyte/monocyte progenitor; RBC, red blood cell. (Adapted, modified and redrawn based on Orkin & Zon, 2008 [2], Kondo *et al.*, 2003 [3] and Proytcheva, 2011 [4])

1.1.1 Ontology of the Haematopoietic System

In mammals, the ontogeny of HSCs occurs sequentially in a variety of distinct anatomical sites [5]. The first wave of haematopoiesis occurs in the yolk sac (YS) blood islands, which

generates large nucleated primitive blood cells expressing embryonic haemoglobin [2]. However, these cells are unable to reconstitute adult long-term haematopoiesis [6]. The next haematopoietic wave (or the definitive haematopoiesis) primarily takes place in aorta–gonad–mesonephros (AGM) region [7] and also other sites including vitelline and umbilical arteries [8] and placenta [9]. Later, these definitive HSCs migrate and colonize foetal liver, thymus, spleen, and finally to the bone marrow (BM), which is the major haematopoietic organ after birth [2]. Unlike primitive haematopoietic system, HSCs derived from AGM are adult-type haematopoietic cells and express adult haemoglobin that proliferate and develop in the haematopoietic niche (BM), and can establish lifelong regeneration of adult blood [10].

The BM is a soft tissue found in between the trabeculae of the cancellous bones [4]. The hollow interior of cancellous bones provides a complex 3 dimensional structure that comprises of endothelial cells, adipocytes, fibroblasts and osteoblasts known as stroma [11]. The BM stromal cells form the microenvironment that mediate direct contact with HSCs, provide extracellular matrix scaffolds, and secrete haematopoietic growth factors that support localization, survival, and self-renewal of HSCs [11]. Studies in mice demonstrate that only a very small proportion (0.05%) of the BM cells is HSCs [12]. These cells belong to a group of multipotent stem cells with self-renewal potential. HSCs are negative for mature lineage markers (Lin$^-$), and express high levels of the stem cell antigen (SCA1$^+$) and the stem cell factor receptor (c-KIT$^+$), and are commonly known as LSK cells [13]. The LSK population is a heterogeneous pool containing 3 subsets of cells: (1) Long term HSCs (LT-HSCs), a stable pool with complete self renewability and capable of long-term multi-lineage haematopoietic reconstitution on lethally irradiated mice [14]; (2) short-term HSCs (ST-HSCs) with limited self-renewal activity, but can rapidly reconstitute and rescue myeloablated transplant recipients [15]; and (3) Multipotent progenitors (MPPs) with no self-renewal capacity, yet retain multilineage differentiation potential [16, 17].

Unlike the HSCs, MPPs have lost self-renewability and begun the journey of irreversible lineage commitment to more differentiated oligopotent progenitors such as common lymphoid progenitor (CLP) and common myeloid progenitor (CMP). However, the ontological branching points of these cells are not fully understood. Several models have been adopted to

explain the process of lineage commitment (**Figure 1.2**). According to the conventional model, further differentiation of MPPs gives rise to either lymphoid restricted [18] or myeloid restricted progenitors [19], which results in a clear separation between lymphopoiesis and myelopoiesis [14]. The alternative model was first proposed by Adolfsson *et al*. in 2001 [20]. They identified a subset of cells known as lymphoid-primed multipotent progenitors (LMPPs) that have lost the potential to differentiate into megakaryocytes and erythrocytes, yet retain the capacity to develop into lymphocytes, granulocytes and macrophages [20, 21].

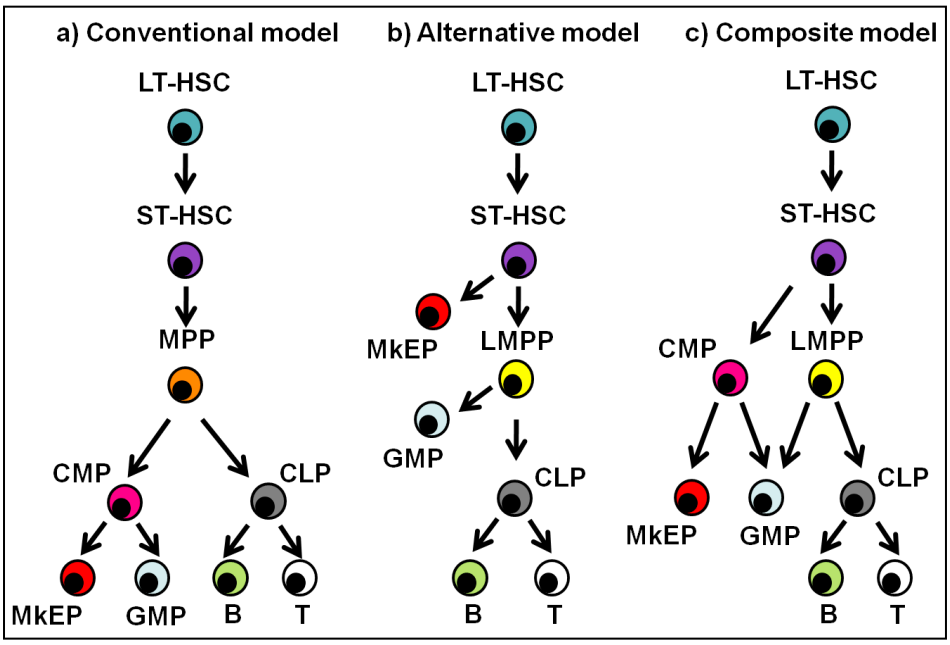


Figure 1.2: Conventional and Alternative Models for Haematopoietic Stem Cell and Blood Lineage Commitment in Mouse.

(A) Conventional model [14] of haematopoietic lineage commitment and development, postulating that the first lineage commitment step of HSCs results in a strict separation of myleopoiesis and lymphopoiesis as supported through the identification of CMPs and CLPs, respectively [18, 19]. (B) Alternative model, based on the present studies, in which a pluripotent HSC, upon loss of MkEP potential, develops into a lymphoid-primed multipotent progenitor (LMPP) that upon loss of GM potential generates the CLP [20]. (C) Composite model, incorporating the experimental evidence for models (A) and (B). LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte/macrophage progenitor; MkEP, megakaryocyte/erythroid progenitor; B, B cell; T, T cell. (Adapted, modified and redrawn based on Adolfsson *et al*. 2005 [21]).

Kondo's group further refined the model and proposed that MPPs generate CMP and GM-lymphoid bipotent lineage during differentiation (**Figure 1.3**) [22]. Subsequently, GM-lymphoid bipotent lineage diverges to lymphoid lineage and a GM differentiation pathway [22]. Taken together, the current data suggest that there is a regression of myeloid development potential as MPPs differentiate towards lymphoid lineage. Collectively, these data bring some important contribution to our understanding of the route for lymphoid and myeloid cell lineages.

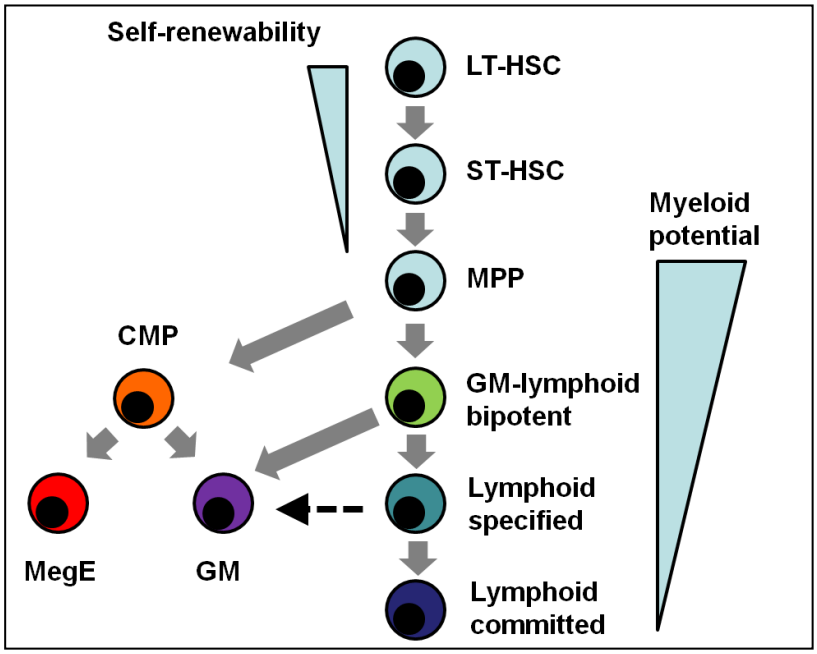


Figure 1.3: A Revised Version of the Classical Model of the Haematopoietic Tree.

A haematopoietic development model proposed by Lai and Kondo [23]. The proposed model suggests CLP and CMP are produced via asymmetrical differentiation of MPPs. MPPs loss myeloid lineage differentiation potential and progress toward lymphoid fate in a step-wise pattern. (Adapted, modified and redrawn based on Lai and Kondo, 2008 [23]).

1.1.2 Transcription Factors in Myelopoiesis

Lineage commitment of haematopoietic cells is governed precisely by a group of transcription factors (TFs). These TFs orchestrate the differentiation programme by driving spatial and temporal expression of lineage specific target genes. Broadly, they can be divided into 2 major groups, the stem cell class of TFs and TFs that are required for multilineage gene expression [2]. In the early days, the field had been ruled by the idea that lineage commitment was

governed by a single lineage restricted TF. However, this idea has changed considerably as numerous studies strongly suggested that combinatorial effects of networks of TFs (cooperate and/or antagonize one another), either as binding partners, co-activators, suppressors or competitors, is required for the commitment of HSC to the fully-fledged mature blood cells [24, 25]. In the next section, I will highlight and discuss several key haematopoietic TFs in myeloid lineage development from HSC.

1.1.2.1 Haematopoietic Stem Cell Transcription Factors

At the stem cell level, stem cell leukaemia/T cell acute lymphocytic leukemia 1 (SCL/TAL1) and runt-related transcription factor 1/acute myeloid leukemia 1 (RUNX1/AML1) are two key haematopoietic TFs. TAL1 belongs to the basic helix-loop-helix (bHLH) family that is required for the primitive (yolk sac blood island) and definitive hematopoietic cell development [26, 27]. Expression of TAL1 is found in the early HSCs and continues to express in the erythroid and megakaryocytic lineage [26, 28]. TAL1 is indispensable in haematopoiesis. Genetic knockout of TAL1 in mice results in complete yolk sac haematopoiesis failure [29], a more severe defect than caused by RUNX1 [30] or GATA2 [31]. Intriguingly, in a conditional knockout model, TAL is dispensable in adult haematopoiesis, suggesting TAL1 is required for the early establishment of the HSC pool [32]. Of note, TAL1 is also expressed in embryonic vascular endothelium, establishing a role for the protein in angiogenesis [33].

RUNX1 (also known as *AML1*, *CBFA2*, or *PEBP2AB*) encodes a DNA-binding subunit of the heterodimeric core-binding factor that is essential for the early stage of definitive haematopoiesis. Developmental stage-specific expression of RUNX1 is evident as it is transiently expressed in definitive haematopoietic cells in emergence sites [34]. Studies have shown that RUNX1 continues to express in adult HSCs and progenitor cells, particularly in myeloid lineages [35]. Knockout of *Runx1* in mouse causes severe haemorrhage, and definitive erythropoiesis and myelopoiesis failure [36]. Similar to TAL1, RUNX1 is not required for maintaining HSCs in the adult. Conditional disruption of *Runx1* inhibited common lymphoid progenitor (CLP) production and blocked maturation of megakaryocytes

and differentiation of T and B lymphocytes [37, 38]. Of note, these mice exhibited marked phenotypes similar to mild myeloproliferative neoplasms (MPN) [37].

GATA2 is another stem cell class TF that is highly expressed in pluripotent HSCs [39]. It is now generally accepted that no single transcription factor drives lineage specific differentiation, but rather combinatorial interaction among TFs. For instance, TAL1, E2A, LDB1, LMO2 and GATA1 form an oligomeric transcription complex in erythroid cells [40]. Interestingly, one study has shown that GATA factors (GATA1/2) are interchangeable in the complex (see **Figure 1.4**). GATA2 predominates over GATA1 during early haematopoiesis. Therefore, GATA2, rather than GATA1 is utilized in the oligomeric protein complex [40]. In this context, GATA2-containing complex regulates expression of HSC maintenance genes such as *c-kit* [41]. Pimanda *et al*. (2007), on the other hand, demonstrated that GATA2, FLI1 and SCL form a fully connected triad of HSC TFs that cross-regulate each other by collectively binding to *Scl*+19, *Fli1*+12 and *Gata2*-3 (HSC enhancers) [42]. Following their initial work, they further investigated combinatorial interactions for ten key regulators (SCL/TAL1, LYL1, LMO2, GATA2, RUNX1, MEIS1, PU.1, ERG, FLI1 and GFI1B) of haematopoietic stem/progenitor cells (HSPCs) [43]. They discovered a new heptad of TFs (SCL, LYL1, LMO2, GATA2, RUNX1, ERG, and FLI1) that regulates HSPCs gene expression in a combinatorial fashion. Of these TFs, 6 bind directly to DNA (all except LMO2) and, depending on the sequence of the regulatory domain, different combinations may interact to stabilise binding and impact on promoter/enhancer efficiency (see **Figure 1.5**). Their studies highlight the important role of the numerous transcriptional networks for cell fate specification of HSCs. In summary, the field of TFs in the stem cell compartment is still in its infancy and the precise role of GATA2 in regulating HSC development is not fully understood. The function of GATA2 is discussed in more detail in section **1.3.1.2**.

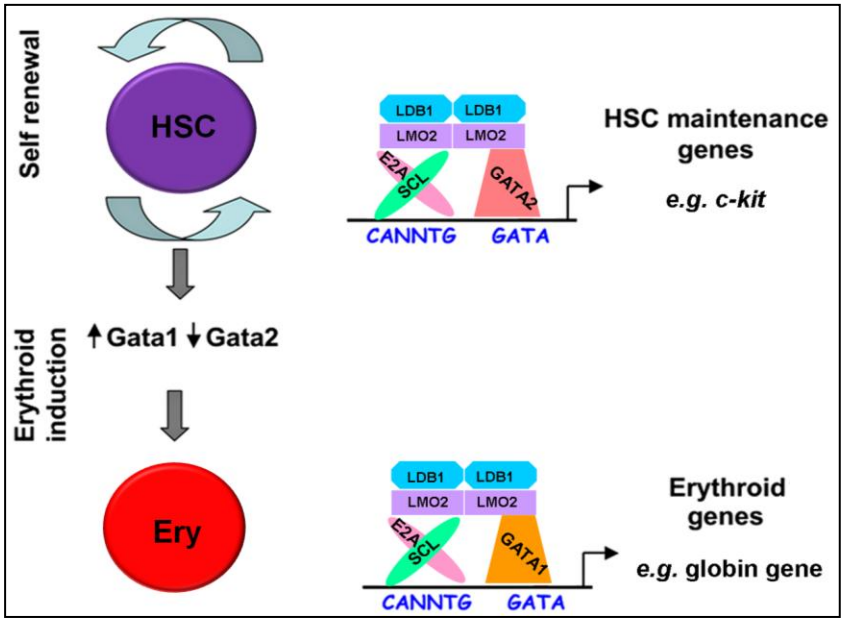


Figure 1.4: Model of GATA2 Function in the Haematopoietic Lineage Determination.

TAL1, E2A, LDB1, LMO2 and GATA1/2 form an oligomeric DNA binding complex in haematopoietic cells. In HSCs, in which GATA2 is highly expressed, LDB1-LMO2-SCL-GATA2 complexes positively regulate expression of HSC maintenance genes. Differentiation of HSCs to the myeloid or lymphoid lineage is trigged by downregulation of LDB1 whereas commitment to the erythroid lineage is triggered by induction of GATA1 and downregulation of GATA2, resulting in the formation of an LDB1-LMO2-SCL-GATA1 complex that positively regulates expression of erythroid-specific genes. (Adapted, modified and redrawn based on Paul Love, 2010 annual report of the Division of Intramural Research, NIH [44]).

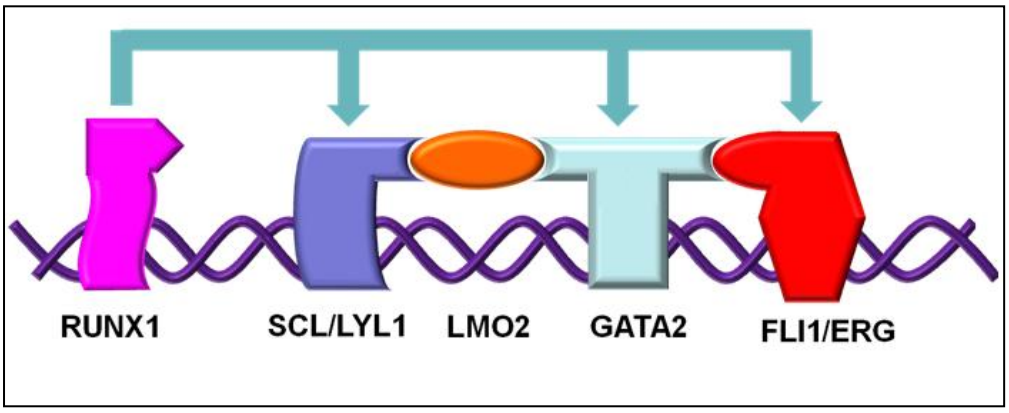


Figure 1.5: Schematic of Heptad Proteins Bound to DNA.

Previously known protein complexes are shown bound to DNA, and the newly identified links between RUNX1 and SCL/GATA2/ERG are indicated by arrows. The order of proteins shown is for illustrative purposes rather than reflecting a particularly common arrangement of binding sites. (Adapted, modified and redrawn based on Wilson *et al.*, 2010 [43]).

1.1.2.2 Common Myeloid Progenitors and Granulocyte Macrophage Progenitors

At the progenitor cell level, several TFs have been shown to play important role in differentiation. Some of these TFs such as PU.1 and CEBPA are known to interact with GATA2. PU.1 (also known as SPI1) is a tissue-specific TF that belongs to the large ETS oncogene family [45]. PU.1 is expressed at varying degrees in HSCs, CMPs and CLPs [46, 47]. For instance, in the myeloid compartment, a low concentration of PU.1 induces granulocyte differentiation, while high levels of PU.1 drive macrophage differentiation [48]. Moreover, multiple studies have shown that PU.1 binds specifically to the promoter region of many myeloid lineage restricted genes [49-51], highlighting PU.1 as a major determinant of myeloid progenitors. Mutations in PU.1 can impair its ability to bind DNA, blocking transcription and differentiation of myeloid progenitors, which is a hallmark of leukaemia [52, 53].

The CCAAT/enhancer binding protein alpha (CEBPA) is a basic-region leucine zipper TF expressed predominantly in immature bipotential myeloid cells [49, 54, 55]. The mRNA of *CEBPA* is translated from two different AUG start codons that give rise to 2 protein isoforms [56]. The CEBPA full-length isoform (p42) has antimitotic activities and promotes differentiation of adipocytes and granulocytes [56, 57]. Conversely, the shorter isoform (p30), which contains the same C-terminus as p42, promotes cellular proliferation [57]. *Cebpa*$^{-/-}$ mice exhibit a complete block in granulopoiesis in foetal and newborn liver, but the development of monocytes and macrophages is not affected [58]. Up-regulation of CEBPA induces granulocytic maturation, but not monocytic differentiation [54, 59]. Conversely, mice with CEBPA conditionally ablated in the haematopoietic system display a phenotype similar to that of AML, as their BM becomes infiltrated by >30% immature myeloblasts [60]. Many studies have demonstrated that 5-14% of *de novo* AML patients contain *CEBPA* mutations [61].

1.1.2.3 Megakaryocyte Erythroid Progenitors

FLI1 and its closely related family member ERG belong the ETS family of TFs that transactivate target genes by binding to the GGA(A/T) core sequence in promoter/enhancer regions [62, 63]. Genetic knockout studies in mice [64, 65] and hemizygous loss of *FLI1* in human [64] have implicated that FLI1 is required for megakaryocytic differentiation.

Furthermore, conditional ablation of FLI1 induces mild thrombocytopenia in adult mice which is associated with a significant reduction of megakaryocytic progenitors and a defect in megakaryocyte differentiation [66]. FLI1 deletion alters lineage commitment switch by driving bipotential MEPs toward the erythrocytic lineage [66], suggesting FLI1 is essential in megakaryocyte maturation. Expression of FLI1 has also been reported in granulocytes, erythrocytes, NK cells and B cells [67, 68]. While GATA1 is a key TF for erythroid and megakaryocytic development, it is also expressed in eosinophils and mast cells. Details for GATA1 are discussed in section **1.3.1.1**.

1.2 Leukaemias and Other Myeloid Malignancies

Leukaemia is one of the ten most common cancers in Australia with 2,591 new cases of reported in 2005. Leukaemia is a prevalent childhood cancer and a major cause of morbidity in adults. Men have at least a 2-fold higher incidence rate compared to women. According to Australian Institute of Health and Welfare (AIHW), leukaemia has an incidence of 12 in 100,000 in the Australian population and the life time risk of developing leukaemia is 1 in 114 by the age of 75 [69]. Despite recent advances in modern medicine, the leukaemia survival rate is still not particularly favourable. A study conducted by the National Statistic Clearing House showed the survival rate of leukaemia patients being 67.1%, 47.8% and 37.9% at 1, 5 and 10 years, respectively. In contrast to the poor survival in adults, survivals in childhood leukaemia have improved significantly over 5 years from 64% in 1982-1986 to 83% in 1998-2004 [70]. The incidence of AML is common in elderly patients with a median age at diagnosis of 67 years old. In contrast, AML in young adults and children is rare and accounts for <7% of total diagnosed AML cases [71]. In the next sections, different aspects of AML including classification, genetics, pathology and aetiology will be discussed.

1.2.1 Classification of Acute Myeloid Leukaemia

Myeloid malignancies represent a group of diseases with very diverse clinical presentations and outcomes. Therefore, a systematic classification is required for diagnosis and management. To date, there are 2 widely adopted classification systems for AML; the older French-

American-British (FAB) system, now largely supplanted by a newer system from the World Health Organization (WHO).

1.2.1.1 French-American-British System (FAB)

The FAB system was first proposed by a group of French, American and British haematologists in 1976 [72]. The objectives of the proposal were to establish an unambiguous nomenclature and uniform classification system for distinguishing myeloid from lymphoblastic leukaemias; and to reduce discrepancies in experimental results and therapeutic trials by different laboratories. In the FAB system, AML is divided into 8 major subtypes (designated as M0 through M7, refer to **Table 1.1**) based predominantly on morphology and cytogenetic abnormalities, to some extent, of the peripheral blood and bone marrow. Each subtype has varying prognoses and responses to therapy. The system appreciates that the presence of blast cells is important in AML and myelodysplastic syndromes (MDS) diagnosis and set a threashold of 30% in peripheral blood (PB) and bone marrow (BM) aspirates for a diagnosis of AML to be considered [73]. The FAB system has been widely used in clinical practice until the WHO introduced a new system in 2001.

Table 1.1: Frequency of FAB Subtypes in a Study Cohort Consists of 614 Adult Patients with *de novo* AML. (Adapted from Haferlach *et al.* 2003 [74])

FAB subtype	Name	Patients (%)
M0	Undifferentiated acute myeloblastic leukaemia	3.3
M1	Acute myeloblastic leukaemia with minimal maturation	19.1
M2	Acute myeloblastic leukaemia with maturation	32.2
M3	Acute promyelocytic leukaemia (t (15:17))	4.9
M3v	M3-variant FAB subtype	1.8
M4	Acute myelomonocytic leukaemia	16.9
M4eos	Acute myelomonocytic leukaemia with eosinophilia	6.2
M5a	Acute monoblastic leukaemia	5.9
M5b	Acute monocytic leukaemia	6.2
M6	Acute erythroid leukaemia	3.3
M7	Acute megakaryocytic leukaemia	0.3

1.2.1.2 WHO Classification

The WHO published a new frame work for haematopoietic myeloid and lymphoid neoplasm classification in 2001 [75]. Under the new frame work, specific haematopoietic and lymphoid neoplasm disease entities are not solely based on morphology and/or cytogenetic information; the WHO criteria also incorporate genetic, immunophenotypic, biologic information as well as clinical features [76]. The new classification aims to provide a mutual consensus among the experts in diagnosis criteria, daily clinical practice as well as for scientific investigations. Following a new revision in 2010, WHO sub-classified AML into 5 major groups and is summarized in **Table 1.2** [77]. The 2010 WHO classification also set the blast count at ≥ 20% as being diagnostic for AML, in contrast to the FAB [75]. Although the WHO system confers more accurate information, the FAB system is still widely used worldwide.

Table 1.2: A Revised WHO Classification of AML. (Adapted from Vardiman 2010 [77])

1) Acute myeloid leukaemia (AML) and related precursor neoplasms
- AML with recurrent genetic abnormalities
  - AML with t(8;21)(q22;q22); RUNX1-RUNX1T1
  - AML with inv(16)(p13.1q22) or t(16;16)(p13.1;p22); CBFB-MYH11
  - Acute promyelocytic leukaemia with t(15;17)(q22;q12); PML-RARA
  - AML with t(9;11)(p22;q23); MLLT3-MLL
  - AML with t(6:9)(p23;q34); DEK-NUP214
  - AML with inv(3)(q21q26.2) or t(3.3)(q21;q26.2); RPN1-EVl1
  - AML (megakaryoblastic) with t(1:22)(p13;q13); RBM15-MKL1
  - AML with mutated NPM1
  - AML with mutated CEBPA

2) AML with myelodysplasia-related changes

3) Therapy-related myeloid neoplasms

4) Acute myeloid leukaemia, NOS (not otherwise specified)
- 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

5) Acute leukaemias of ambiguous lineage
- Acute undifferentiated leukaemia
- Mixed phenotype acute leukaemia with t(9;22)(q34;q11.2); BCR-ABL1
- Mixed phenotype acute leukaemia with t(v;11q23); MLL rearranged
- Mixed phenotype acute leukaemia, B/myeloid, NOS
- Mixed phenotype acute leukaemia, T/myeloid, NOS
- Natural killer (NK) cell lymphoblastic leukaemia/lymphoma

1.2.2 The Genetics of Myeloid Malignancies

Haematopoiesis is a stepwise differentiation process of HSCs into mature blood cells with specific functions. Haematopoietic perturbations by intrinsic factors (TFs) and/or extrinsic factors (growth factors) may endow a proliferative and/or survival advantage and an impaired differentiation potential to immature precursors, leading to loss of lineage identity, increased

infidelity, lineage reprogramming, and ultimately, malignant transformation [78]. In the myeloid compartment (also in lymphoid cells), two common genetic lesions have been described: (1) submicroscopic gene mutations that alter its function (*e.g.* LOF and GOF) (2) microscopic visible structural and numerical chromosome alterations [79]. In recent years, extensive experimental evidence has shown that epigenetic modifications and small regulatory RNAs (microRNAs) also play important roles in AML [80].

1.2.2.1 Intermediate and Small Genetic Lesions

Intermediate genetic lesions refer to deletions and insertions between 10-3,000bp that are not easily detectable by whole exome sequencing, comparative genomic hybridization, SNP arrays or cytogenetics. These may be found by whole genome sequencing, but are still difficult to detect. Small genetic lesions, on the other hand, refer to point mutations and small deletions or insertions. Both intermediate and small genetic lesions often occur in cytogenetically normal AML (CN-AML). In fact, CN-AML accounts for 40-50% of all AML cases in patients [81]. These CN-AML patients are associated with an "intermediate" prognosis. Molecular analyses of CN-AML samples have identified some important small genetic lesions that target TFs.

RUNX1 is a TF that is associated as a fusion protein with other genes (see Atlas of Genetics and Cytogenetics in Oncology and Haematology [82]) and hence predominantly in karyotypically abnormal AML [83] (see section **1.2.2.2**). However, it also commonly acts as a tumour suppressor gene in CN-AML with disrupting mutations in both *RUNX1* alleles. Molecular analysis reveals *RUNX1* point mutations account for 10% of sporadic CN-AML [84], predominantly in FAB subtype M0 with 56% of them being biallelic mutations [85]. Germline *RUNX1* mutation (heterozygote) predisposes affected individuals to familial platelet disorder with increased risk to develop AML (FPD/AML) [86]. Genetic lesions of *RUNX1* have also been reported in MDS and MDS/AML [87]. Intriguingly, Harada *et al*. described that *RUNX1* point mutations (frame-shift/nonsense mutations) in sporadic MDS and MDS/AML are predominantly in the C-terminal region (*trans*-activation domain), suggesting mutation in this region is likely a distinct feature of MDS/AML [87]. Most of the *RUNX1* small genetic lesions are missense mutations that result in LOF [85]. Others are nonsense mutations that produce truncated proteins [85]. Importantly, the above mentioned studies

showed that a reduction of RUNX1 activity, resulting from haploinsufficiency is associated with increased susceptibility to myeloid malignancies [85].

CEBPA contributes to approximately 9-13% of sporadic CN-AML [88, 89]. Like *RUNX1*, germline mutations have also been reported in several families that predispose to AML [90, 91]. Pabst *et al*. first described heterozygous frameshift mutations in *CEBPA* that disrupt production of p42 (isoform 1), but preserve and enhance production of p30 (isoform 2), and produce a truncated protein (short N-terminal, p20). Both p30 and p20, however, inhibit WT CEBPA (p42) DNA binding and transactivation of granulocyte target genes and block granulocytic differentiation [92]. Subsequently, others have discovered insertion and deletion mutations that disrupt the basic-Leucine-Zipper (bZIP) domain, resulting in LOF of CEBPA proteins [93, 94]. Nearly 50% (28/59) of the CEBPA mutated patients harbour either N-terminal frameshift mutations or C-terminal inframe indel mutations in their genome [95]. Biallelic *CEBPA* mutations are sometimes found in both sporadic and familial AML patients [96]. These patients harbour a mutation in bZIP domain in one allele and a N-terminal frameshift mutation in the other copy [57]. An important question is what are the differences between monoallelic and biallelic *CEBPA* AMLs in term of phenotype? In murine models, biallelic mutation was shown to accelerate AML development [97]. A matching result was demonstrated by Kato *et al*. in which they discovered that monoallelic mutation is predominantly in MDS/AML, while biallelic is more common in *de novo* AML, suggesting these mutations may have different mechanisms of leukaemogenesis [98].

1.2.2.2 Large Genetic Lesions

Large genetic lesions refer to chromosomal abnormalities such gain or loss of whole chromosome, chromosomal deletion (partial), aneuploidy, chromosomal translocation, duplication and inversion. All are frequently associated with tumourigenic potential. In fact, approximately 50-60% of adult AML have cytogenetic abnormalities [99] and some of the most common translocations involve RUNX1 or its co-factor (CBFB). For instance, the *RUNX1-ETO* t(8;21) translocation is one of the most common recurrent genetic lesions found in ~5-10% of all *de novo* AML cases and 10-22% of FAB subtype M2 cases [100]. The inversion of chromosome 16 [inv (16) (p13q22)] (also known as CBFB-MYH11) is another

common chromosomal abnormality associated with acute myelomonocytic leukemia (M4 subtype) with eosinophilia [101]. Nevertheless, AML patients harbouring *RUNX1-ETO* or *CBFB-MYH11* are considered to have a good prognosis and are likely to respond well to therapy.

1.2.3 Acute Myeloid Leukaemia and Myeloid Neoplasms

AML is a heterogeneous disorder with a myriad of genetic lesions arising from neoplastic HSC or progenitor cells. The disease is characterized by uncontrollable proliferation and loss of differentiation of myeloid cells. The result is the accumulation of blast cells (immature, non-functional and undifferentiated precursor cells), which inhibits normal haematopoiesis. If left untreated, AML is universally fatal. Clinical presentations of AML are commonly nonspecific, but they result from marrow failure and the subsequent cytopaenias [102]. This includes fatigue (from anaemia), easily bruising and bleeding with minor trauma (from thrombocytopaenia) and susceptibility to infections (neutropaenia).

Recent advances in molecular biology and genomic technologies have led to the identification of a myriad of genetic lesions in AML. In many AML cases, the primary lesions are oncogenic fusion proteins, which are the end result of chromosome translocations [103]. The second key feature in AML is genetic lesions caused by mutations. Briefly, these lesions often fall into several major groups including: (1) abnormalities in the cytokine networks, (2) mutations in growth factor receptors, (3) activating mutations in kinase-mediated signalling cascades, (4) mutations in key haematopoietic transcription factors in differentiation and (5) defects in cell cycle checkpoints and epigenetic alterations (**Figure 1.6**) [104]. Together, all these genetic aberrations confer a proliferative and/or survival advantage and impairment of differentiation, leading to leukaemogenesis [104]. Based on the clinical presentations, AML can be conceptually sub-classified into *de novo* AML (or primary) and secondary AML [105]. The latter class may be further sub-divided into: (1) patients with antecedent haematological disorders such as myelodisplastic syndromes (MDS) or myeloproliferative neoplasms (MPN) and (2) therapy-related AML (t-AML), in which patients usually have genetic damage from prior radiotherapy or chemotherapy for preceding cancers.

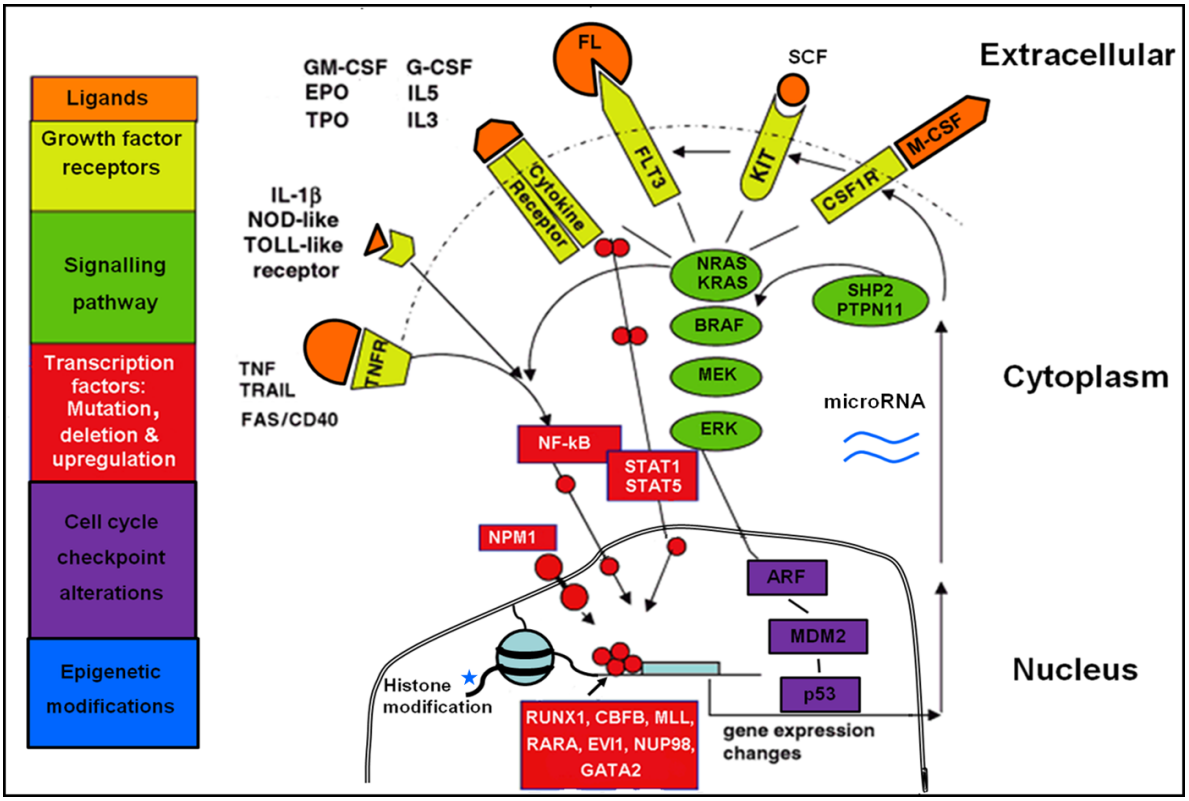


Figure 1.6: AML-Related Genetic Pathways.

A general schematic depicting the heterogeneity in the genetic lesions associated with AML. An abnormal production of haematopoietic growth factors and mutations of their receptors may lead to AML via a dysfunction autocrine growth mechanism [106, 107]. Other common mutations found in AML (also MDS) patients are (1) Activation of class I mutation such as tyrosine kinase-RAS/BRAF mediated signalling pathway that leads to cell proliferation; (2) Perturbation and deregulation (suppression or inactivation) of key haematopoietic transcription factors (class II mutations) that results in perturbation of cell differentiation; and (3) Mutations of the cell cycle checkpoint related proteins such TP53, ARF and MDM2 [107, 108]. Recently works have established that epigenetic alterations such as histone modification and small regulatory RNAs (microRNAs) also play important roles in leukaemogenesis (Picture adapted, modified and redrawn based on Dan Jones, 2010 [104]).

1.2.3.1 *De novo* AML

Unlike secondary AML, *de novo* AML occurs in patients with no preceding haematological disorders or prior exposure to carcinogens. Recently discovered information using cytogenetics and molecular genetics allow *de novo* and secondary AML to be differentiated as biologically distinct entities [109, 110]. In comparison to *de novo* AML, secondary AML patients (*e.g.* t-AML) are generally more resistant to chemotherapy and have a poor long term

survival [111-113], whereas, *de novo* AML is usually a rapid onset, clinically aggressive disorder with a high initial complete remission rate after chemotherapy [114, 115]. However, it should be emphasized that treatment outcomes may depend on specific chromosome aberrations. For example, the presence of t(8;21), t(15;17) and inv(16)/t(16;16) is generally recognized as good prognosis in *de novo* AML [116, 117]. In contrast, the presence of -5, 5q-, -7, 7q-, and +8 is associated with poor prognosis regardless of whether it is in *de novo* AML or secondary AML [105].

For many years, chromosome abnormalities have been recognized as one the hallmarks of haematological malignancies. Overall, chromosome aberrations are often balanced in *de novo* AML via reciprocal chromosomal translocations or inversion rearrangements without visible gain or loss of DNA materials [114, 118]. In support of this concept, Mauritzson *et al*. (2002) demonstrated that complex karyotypes and hypodiploidy are more common in t-AML and therapy-related myelodysplastic syndromes (t-MDS) than their *de novo* counterparts in a pooled analysis study [109]. Intriguingly, in elderly *de novo* AML patients, cytogenetic studies demonstrated karyotypic abnormalities resembling secondary AML (*i.e.* unbalanced chromosome translocation and complex karyotype, with gain and/or loss of multiple chromosomal regions) [119, 120].

1.2.3.2 Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS), also previously known as “pre-leukaemia”, are a clonal stem cell disorder characterized by ineffective haematopoiesis in one or more of the lineages of the bone marrow which have a high propensity to evolve to AML (or MDS/AML) [121]. Almost 50% of the MDS patients are asymptomatic during the time of initial diagnosis [122]. As the disease progresses, most MDS patients present anaemia, neutropaenia and/or thrombocytopaenia, leading to fatigue, infection, bleeding and other symptoms [122, 123]. MDS in children and young patients are relatively rare; the risk of MDS increases with age and often has an onset over 60 years old [124].

MDS and AML are complex and heterogeneous diseases with overlapping features and they can be difficult to distinguish from each other on clinical grounds in some cases. They both

manifest with cytopaenia and similar BM changes. By definition, ≥20% blasts is diagnostic of AML, though in many cases of MDS, the blast count is close to 20%, especially in late disease. They also share some recurring cytogenetic abnormalities such as chromosome 5/7 deletions, though the prognostic significance can be different in each disease. For instance, in MDS, 5q- is a good prognostic factor, whereas it is a poor risk marker in AML [121, 125].

Progression of MDS to AML (MDS/AML) is a multistep process of accumulating genetic mutation events. Longitudinal studies have shown that a broad spectrum of genetic lesions including chromosomal aberrations, gene mutations, epigenetic changes such as DNA hypermethylation and histone deacetylation, changes in the BM microenvironment and dysregulated immune surveillance are associated with the leukaemic transformation [126]. Nearly 30% of patients with high risk MDS progress into AML, thus, MDS/AML is a good model for studying the multistep process of leukaemogenesis [127].

1.2.3.3 Myeloproliferative Neoplasms

Myeloproliferative neoplasms (MPN) represent a group of phenotypically diverse diseases that characterized by excessive production of one or more terminally differentiated myeloid cell lineages in the peripheral blood and have an increased long term risk to transform into AML [128, 129]. MPNs were first described by William Dameshek in 1951 by grouping together polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF) and chronic myeloid leukaemia (CML) [130]. CML is characterized by a reciprocal translocation between chromosomes 9 and 22 t(9;22) (q34;q11) which generates two fusion genes, BCR-ABL on the Philadelphia chromosome (Ph) and ABL-BCR on 9q+ [131, 132]. BCR-ABL has constitutive tyrosine kinase activity causing unregulated growth [133], which is regarded as the central hallmark molecular mechanism underlying CML pathogenesis [134]. Moreover, Ph^+ clones are prone to further genetic aberrations which drive chronic phase (CP) to accelerated phase (CML-AP) and ultimately to blastic phase (CML-BP) when left untreated. Mutations and dysregulation of several critical genes, notably loss of p53 [135], p16 [136] and pRB [137], and a GATA2 L359V gain-of-function mutation [138], have been shown to contribute to malignant progression to CML-BP, a cancer that is akin to AML in many ways.

Similarly, three other Philadelphia chromosome negative MPN (PV, ET and PMF) may progress and transform to AML (secondary AML) [139-142].

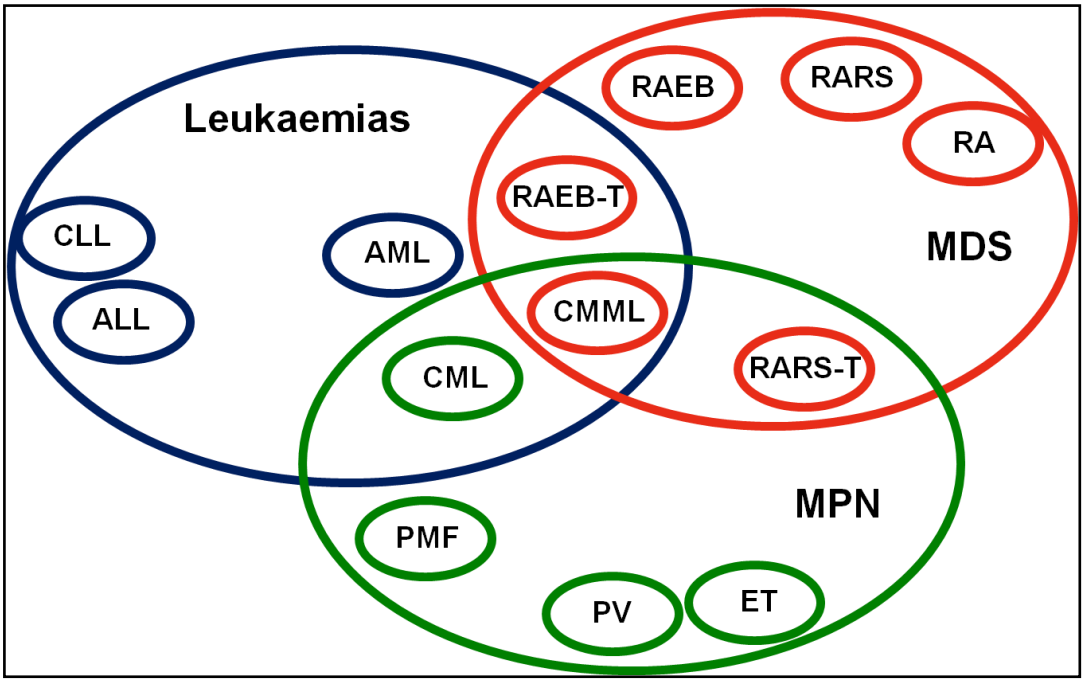


Figure 1.7: The Clonal Bone Marrow Disorders.

Patients with MDS or MPN have an increase risk of transformation into acute myeloid leukaemia. According to the FAB system, early MDS including refractory anaemia (RA) and refractory anaemia with ringed sideroblasts (RS) are associated with an increase in apoptosis and a reduction in proliferation. During disease progression, RA/RARS may evolve to refractory anaemia with excessive blast (RAEB), and ultimately progress to the advanced stage known as refractory anaemia with excessive blast in transformation (RAEB-T). At this stage, a significant reduction in apoptosis is observed [125, 143]. A subset of patients develops RARS with marked thrombocytosis (RARS-T), a MDS/MPN overlap syndrome that exhibits the features of both RARS (classified with MDS) and thrombocytosis (feature suggestive of MPN) [144].

PV and ET may progress to secondary myelofibrosis (MF), which clinically resembles PMF. Rarely, PV and ET patients progress to MPN-BP without going through a phase of MF (primary or secondary MF) [128]. PMF and post-PV/ET MF patients are predisposed to leukaemic transformation or MPN-BP. On the other hand, CML represents the chronic phase of myeloid leukaemia that may progress to AML with additional mutation hits. CMML (chronic myelomonocytic leukaemia) has been revised as myelodysplastic/myeloproliferative neoplasm (MDS/MPN), a disease which comprises mainly mature cells with both effective and ineffective proliferation of various lineages [77]. (Figure adapted, modified and redrawn based on Kouides & Bennett, 1996 [145]).

1.2.4 Pure Familial Leukaemia

The concept of pure familial leukaemia was initially discussed by Horwitz to distinguish true high risk families from others that have more a complex medical syndrome (*e.g.* Fanconi anaemia and Bloom syndrome, see section **1.2.5**) in which offspring inherit a constellation of abnormalities, one of which is predisposition to leukaemia [146]. The familial aggregation of pure myelodysplastic syndrome (MDS) and/or acute myeloid leukaemia (AML) is relatively rare. At the commencement of my studies, the only well described heritable mutations were in *RUNX1* and *CEBPA*. Mutation of these genes results in a Mendelian disorder with variable penetrance, typically with an autosomal dominant inheritance fashion [147]. Mutations in these genes are also commonly seen in sporadic AML [87, 148] and studying rare familial leukaemia can provide a gateway to identify causative genes in the more common sporadic cases.

1.2.4.1 *RUNX1* Associated Familial Platelet Disorder with Propensity to AML

Rare familial aggregations of AML have been described in multiple studies. For instance, familial platelet disorder with a propensity to develop AML (FPD/AML; OMIM: 601399) is an autosomal dominant inherited disease characterized by thrombocytopaenia, abnormal platelet function and a lifelong risk of development of haematologic malignancies [86]. FPD/AML was first reported by Dowton *et al*. (1985) in a large French-Canadian pedigree that consisted of more than 20 affected family members [149]. Ho *et al*. (1996) reported the genetic linkage of the FPD/AML markers on chromosome 21q22 [150]. Subsequently, Song *et al*. (1999) identified a germline heterozygous mutation in *RUNX1* [86]. Of note, somatic *RUNX1* mutations are quite common in MDS and AML [87, 148].

1.2.4.2 *CEBPA* Associated Leukaemia

Smith *et al*. (2004) reported the first germline *CEBPA* mutation to cause dominant inheritance of AML (OMIM: 601626) in three family members over two generations [91]. Subsequently, 2 other families were reported [90, 151]. Unlike FPD/AML with variable clinical and pathological phenotypes, *CEBPA* mutated patients present consistent features including FAB M1 or M2 subtypes, many Auer rods and aberrant CD7 expression on myelobalsts and normal karyotypes in both germline and sporadic cases [96]. Patients with *CEBPA* mutation are

generally considered to have a good prognosis in both germline and sporadic AML patients [96], although not in all studies [152].

1.2.4.3 *GATA2* Associated Myeloid Malignancies

We first reported the discovery of *GATA2* mutations in 4 families that are susceptible to MDS/AML (3 large families; OMIM: 601626) and MDS (1 small family; OMIM: 614286) (This finding is a major part of the work in this thesis; see **Chapter 3** and Hahn *et al.* (2011) [153]). Molecular analysis on MDS/AML families revealed a germline transmission of a heterozygous *GATA2* missense mutation (T354M; OMIM: 137295.0002) that segregated with disease. In the MDS family, a heterozygous 3-bp deletion in *GATA2* (T355del, OMIM: 137295.0014) was identified in father and son. We also performed thorough functional studies on T354M, T355del and L359V. Details of the discovery are outlined in **Chapter 3**.

1.2.5 Hereditary Syndromes with Predisposition to Leukaemia

Haematological malignancies may develop as a consequence of an inherited syndrome (as opposed to pure familial AML). Several inherited genetic syndromes have long been recognized as one of the possible risk factors for the development of haematological malignancies (**Table 1.3**), though they primarily manifest other disease characteristics [154]. For instance, patients with DNA repair syndromes are often prone to develop cancers, particularly haematological malignancies. Similarly, a subset of patients with antecedent bone marrow disorders such as severe congenital neutropaenia, Schwann-Diamond syndrome and Diamond-Blackfan syndrome, progress to leukaemias. Autosomal dominant inheritance of Li-Fraumeni syndrome and Neurofibromatosis greatly increase susceptibility to cancers including leukaemias. Individuals with inherited syndromes associated with immunodeficiency such as severe combined immunodeficiency, Wiskott-Aldrich syndrome, X-linked immunodeficiency and X-linked lymphoproliferative, also have an increased risk for developing haematological malignancies. Lastly, dyskeratosis congenita caused by an autosomal dominant mutation leading to telomere related defects predispose to AML [154]. In summary, an insight into the oncogenic pathways conferred by these genes could lead to a better understanding of leukaemogenesis.

Table 1.3: Genetic Factors Predisposing to the Development of Secondary AML.

Abbreviations: AR, autosomal recessive; AD, autosomal dominant; XR, X-linked recessive, T-PLL, T-cell-prolymphocytic leukaemia; AMML, adult myelomonocytic leukaemia; JMML, juvenile myelomonocytic leukaemia; EL, erythroid leukaemia; AMoL, acute myelomonocytic leukaemia; BL, Burkitt lymphoma; HL, Hodgkin lymphoma and AA, aplastic anaemia. (Table adapted from Carmichael & Scott, 2007 [154]. Data were taken from Sankaran *et al*. 2012[#] [155] and Yamaguchi *et al*. 2003* [156]).

Syndrome	Gene	Inheritance	Haematopoietic malignancies	Approximate risk
DNA repair syndromes				
Ataxia telengiectasia	*ATM*	AR	T-cell lymphoma, T-ALL, T-PLL, B-cell lymphoma	12%
Bloom syndrome	*BLM*	AR	AML, ALL, lymphoma	35%
Fanconi anaemia	*FANC*	AR	AML	10%
Nijmegan breakage	*NBS1*	AR	Lymphoid	Unknown
Bone marrow failure syndromes				
Severe congenital neutropaenia/ Kostmann syndrome	*ELA2*, *GFI1*, *HAX1*	AR, AD	AML, AMoL	2-10%
Schwann-Diamond	*SBDS*	AR	ALL, AML, AMML, AMoL, EL, JMML	5%
Diamond-Blackfan	Linkage to 19q or 8p (80%) or *RPS10* (20-25%) *GATA1*[#]	AR and AD	AML	4%
Tumour suppressor syndromes				
Li-Fraumeni	*p53*	AD	B-CLL, ALL, CML, HL, BL, JMML, AML	50% for all cancers
Neurofibromatosis	*NF1*	AD	JMML, AML	350-fold increase
Immunodeficiency syndromes				
Severe combined immunodeficiency	*ADA*	AR	B-cell lymphoma	5%
Wiskott-Aldrich	*WASP*	XR	ALL, HL	7%
X-linked immunodeficiency	*CD40L*	XR	Lymphoma, HL	Unknown
X-linked lymphoproliferative	*SAP*	XR	EBV-related B-cell lymphoma	20%
Telomere syndrome				
Dyskeratosis congenita*	*TERC*	AD	AA, MDS	Unknown

1.2.5.1 Emberger Syndrome

Emberger syndrome is a rare autosomal dominant disorder (OMIM: 614038) characterised by sensorineural deaf mutism, lymphoedema of lower limbs with early onset MDS associated with an increased risk of AML transformation [157]. The disease was initially investigated by several groups in families affected with the disease [157-159].

Mansour *et al*. (2010) reported 7 unrelated Emberger syndrome cases [160]. They noted that these patients exhibited early onset of lymphoedema (infancy and puberty) that affected one or both lower limbs and the genitalia prior to onset of haematological abnormalities. Other notable features in these patients include hypotelorism, epicanthic folds, long tapering fingers and/or neck webbing (four patients), recurrent cellulitis in the affected limb (four patients), generalized warts (two patients), and congenital, high frequency sensorineural deafness (one patient). We initially found lymphoedema in a patient with a large deletion (Chromosome 3) encompassing GATA2 and surmised that haploinsufficiency of GATA2 was responsible for the lymphatic phenotype because of the high expression of GATA2 in lymphatic endothelial cells [161]. Subsequently, Mansour's group performed whole exome sequencing on 3 unrelated patients with Emberger syndrome (2 familial cases from their previous study and 1 new sporadic case) and discovered 3 different *GATA2* heterozygous frameshift mutations (OMIM: 137295.0009-137295.0011) [162]. Germline transmission of *GATA2* mutation was also confirmed in the 2 familial cases. They further sequenced 5 sporadic individuals (3 were previously reported by Mansour *et al*., 2010) and identified 2 novel *GATA2* heterozygous missense mutations, C373R (OMIM: 137295.0012) and R361L (OMIM: 137295.0013) in 2 patients (others are frameshift mutations). In summary, their study highlights that *GATA2* is the predisposition gene for Emberger syndrome.

1.2.5.2 MonoMac/DCML Deficiency Syndrome

MonoMac (monocytopaenia with *Mycobacterium avium* complex) or DCML (dendritic cell, monocyte, B and natural killer (NK) lymphocytes) deficiency syndrome (OMIM: 614172) is a disease characterized by significantly decreased or absence of monocytes, lymphocytes, and circulating and tissue dendritic cells (DCs) with variable T cell lymphocytopaenia [163, 164]. Clinical presentations of MonoMac/DCML deficiency syndrome include susceptibility to

disseminated nontuberculous mycobacterial infections, viral infections, especially with human papillomaviruses, and fungal infections, primarily histoplasmosis, and pulmonary alveolar proteinosis [163, 164]. Clinical presentations of MonoMac/DCML deficiency syndrome include susceptibility to disseminated nontuberculous mycobacterial infections, viral infections, especially with human papillomaviruses, and fungal infections, primarily histoplasmosis, and pulmonary alveolar proteinosis [164]. Bigley *et al*. (2011) described 4 unrelated patients (3 sporadic and 1 familial) with severe depletion of monocytes, B and NK lymphocytes and DCs in their peripheral blood [163]. Both autosomal dominant transmission and sporadic cases have been reported in these patients [163, 164]. Neither of them identified disease causing genes or predisposition genes.

Recognizing a subset of MonoMac patients presented a syndrome similar to those that reported by our group (Scott *et al*. 2010) [165], Hsu *et al*. performed targeted sequencing on 20 patients [166]. 12 distinct *GATA2* mutations (mostly in ZF2 domain) were indentified in 10/16 kindreds that were previously reported by Vinh *et al*. (2010) and in 8 new kindreds. Of note, 2 recurrent mutations, T354M (3 cases; OMIM: 137295.0002) and R398W (5 cases; OMIM: 137295.0001) were discovered in unrelated patients. Other mutations included 6 patients with insertion/deletion in *GATA2* that results in either frameshifts or premature termination or null alleles, 4 missense mutations and a mutation of a canonical splice acceptor. In the same line of reasoning, Dickinson *et al*. (2011) re-investigated 4 DCML deficiency patients that were previously reported by Bigley *et al*. (2011) [167]. They identified 4 *GATA2* mutations which included 2 recurrent mutations (T354M and R398W) and 2 novel mutations (G200fs and Δ340-381). The framseshift and deletion mutations were thought to cause complete loss of protein function, resulting in haploinsufficiency of GATA2. However, no GATA2 functional studies were performed.

1.3 Introduction to GATA Family of Transcription Factors

The mechanism of haematopoiesis is one of the most extensively studied systems in which pluripotent HSC undergo an irreversible maturation process and differentiate into lineage-

specific functioning progenitor cells (see section **1.1**). Cell fate commitment begins with a cascade of differentially expressed TFs that dictates a particular cell fate. Haematopoietic GATA factors, particularly GATA1, 2 and 3 are prototypical examples in promoting haematopoiesis. In this section, the haematopoietic GATA family members and their specific roles are discussed.

1.3.1 GATA Family Members

To date, there are 6 GATA family transcription factors have been identified in vertebrates. Each of the members shares designated features, yet exhibits unique functional properties. Nevertheless, all GATA binding proteins share 3 common features: (1) These transcription factors bind to a core DNA consensus sequence (A/T)GATA(A/G) in the regulatory regions; (2) they contain 2 zinc finger domains which are conserved among the family members and the regions other than conserved domains have low homology [168]; (3) Expression of the GATA family members are tissue specific. GATA1, 2 and 3 are essential, but play distinctive roles (albeit not exclusively) in haematopoietic development [169]. Expression of GATA1 is found in differentiated cells of the erythroid, megakaryocytic, mast and eosinophilic lineages; suggesting that GATA1 is essential in erythroid and megakaryocyte differentiation [170]. GATA2 functions in haematopoiesis and is required for proliferation, survival and homeostasis of immature haematopoietic progenitors, and mast cell formation [171]. GATA3 is indispensable in T-cell development [172]. In contrast, GATA4, 5 and 6 belong to sub-family members that are expressed widely in various mesoderm- and endoderm-derived tissues such as heart, liver, lung, gonad, and gut [173].

1.3.1.1 GATA1

GATA1 is a chromosome X-linked gene [174] that initially identified as a trans-acting factor that specifically regulates promoter and enhancer regions of α- and β-globin gene clusters [175, 176]. With the exception of the multipotential progenitors, high expression levels of GATA1 have been reported in primitive and definitive erythrocytes [177, 178], megakaryocytes [179], mast cells [179], eosinophilic cells [180] and in the Sertoli cells of the testis [181, 182]. Studies of GATA1 provide numerous insights into the importance of GATA1 function in these cells.

GATA1 is absolutely vital for normal erythropoiesis. *Gata1* disrupted embryonic stem (ES) cells failed to generate mature erythroblast [183, 184] and underwent death by apoptosis [185, 186]. Concordantly, Takahashi *et al*. have shown that conditional *Gata1* promoter knock-out male mice developed severe anaemia and died *in utero* at 12.5 days; while heterozygous females survived and displayed varying degrees of anaemia to normal erythropoesis [187]. The surviving animals developed normally, yet their lifespan was significantly shorter than wild-type littermates. Post-mortem examination revealed these mutant mice displayed marked splenomegaly, anaemia and thrombocytopenia [188]. Nevertheless, the embryonic lethality phenotype was fully rescued by transgenic expression of *Gata1* cDNA. Interestingly, GATA2 and GATA3 could also partially rescue the GATA1 deficiency phenotype, suggesting that the haematopoietic GATA factor proteins are partially redundant, but not functionally equivalent [189].

GATA1 also plays a critical role in megakaryopoiesis. How GATA1 orchestrates megakaryocyte development has been elucidated by multiple lines of evidence. Firstly, forced expression of GATA1 drives murine myeloid cells 416B to megakaryocytic differentiation [190]. Secondly, exogenously overexpressed GATA1 reprograms avian Myb-Est-transformed myeloblasts to differentiate into thromboblasts in a dose-dependent manner [191]. Thirdly, *Gata1* null foetal liver cells derived from chimeric mice display a 4-fold increase in the number of mature megakaryocytes, implicating that GATA1 affects the balance and kinetics of megakaryocyte formation and turnover [183]. Moreover, GATA1 deficient megakaryocytes possess a growth advantage over their WT counterparts *in vitro* [192]. Lastly, *Gata1* mutant mice exhibit increased proliferation and impaired maturation of megakaryocytic progenitors, and the animals suffer from consistently lower platelet counts [188, 192], although full maturation was achievable in another model [183]. The mutant mice produce defective platelets and are not fully functional due in part to abnormal structure [192]. Analysis of *in vitro* and *in vivo* studies provides direct evidence that GATA1 is a key regulator of megakaryocytic lineage maturation.

The first direct evidence of involvement of GATA1 in eosinophil maturation was reported by Kulessa and co-workers, in a lineage reprogramming study via ectopic overexpression of

GATA1 in chicken Myb-Est-transformed myeloblasts, which induced differentiation of these cells into eosinophils [191]. Subsequently, Hirasawa indicated that enforced expression of GATA1 instructs CD34 positive haematopoietic progenitor cells to develop into eosinophils. Moreover, GATA1-deficient mice failed to develop eosinophil progenitors in the foetal liver [193]. Remarkably, the deletion of a double GATA binding site in the *Gata1* promoter incurs selective eosinophil lineage depletion *in vivo* [194], implicating the palindromic GATA site is required for autoregulation of *Gata1* [195].

The importance of a role of GATA1 in mast cell differentiation and activation has been evidenced by several investigators [67, 179, 196]. Distinct from other haematopoietic cells, mast cell precursors circulate in the blood stream and only mature once in connective or mucosal tissues. GATA1 is more highly expressed in differentiated BM-derived mast cells than BM progenitors, implicating the protein is required for mast cell terminal differentiation [197]. Furthermore, one study has shown reduced expression GATA1 can impair mast cell differentiation *in vivo* [198]. However, the phenotypes are reversible using retroviral-mediated expression of *Gata1* cDNA [198].

1.3.1.2 GATA2

GATA2, the principal focus of this thesis, was first isolated from a chicken cDNA library [199]. Over the last 2 decades, extensive experimental data of the role of GATA2 has been well documented. In some cells, the GATA2 expression profile overlaps with GATA1, particularly in haematopoietic lineage cells. In these lineages, GATA2 plays a pivotal functional role in mast cells [196], eosinophils, neutrophils and basophils [180], early erythroid differentiation [178], megakaryocyte differentiation [200], pluripotent hematopoietic stem cells [39], early haematopoietic progenitor cells [201], and in the immature haematopoietic cells derived from the aorta-gonad-mesonephros (AGM) region and foetal liver [202]. Outside of haematopoiesis, GATA2 expression and function have been reported in androgen regulation [203], inner ear development [204], midbrain GABAergic neurons [205], adipocyte differentiation [206] and in the vascular endothelium [207].

GATA2 expression is controlled by IS and IG promoters. The distal first exon (IS) is utilized by SCA-1^{+}/c-KIT^{+} haematopoietic progenitor cells and neural cells. In contrast, the mRNA that initiates with proximal IG first exon is expressed more widely in various tissues [208, 209]. To further dissect GATA2 IS promoter function, Minegishi *et al*. knocked a green fluorescence protein (GFP) into the regulatory regions of the IS exon [210]. The transgenic reporter was uniquely expressed in the early haematopoietic tissues and were able to generate both immature haematopoietic and endothelial cell clusters *in vitro* [210], indicating that IS promoter is utilized in these cells. Most strikingly, these homozygous IS exon GFP knock-in mice were viable and fertile. Indeed, their GATA2 expression level was similar to that of WT littermates, leading to a hypothesis that the IG promoter compensates for the IS promoter activity in these mice in these cells [211].

GATA2 deficiency leads to an early lethal phenotype and *Gata2*$^{-/-}$ mice died at embryonic day 10 to 11 with severe anaemia [212]. Consistent with this result, *Gata2*$^{+/-}$ mice exhibit a significant reduction in the production, self-renewal and expansion of HSCs in the AGM, but not in yolk sac, foetal liver or adult BM, which showed minimal changes [213]. In striking contrast to the normal mice, *Gata2*$^{+/-}$ mice have a reduced number of HSCs which was associated with increased apoptosis and cellular quiescence, while the self-renewal and differentiation capacity of HSCs was unaffected [214]. Furthermore, the committed granulocyte-macrophage progenitor (GMP) lineage derived from *Gata2*$^{+/-}$ BM was attenuated in number and displayed impaired functionality *in vitro* and *in vivo* [215]. However, *Gata2*$^{+/-}$ had little or no effect on CMPs and CLPs [215]. Together, these observations also lead to the postulation that GATA2 plays a stage-specific differentiation role in the HSC and GMP compartments separately. To circumvent the embryonic lethality problem, conditional knockout mice were generated. Charles *et al*. reported a conditional knockout strategy by using the site-specific recombinase system *Cre-LoxP* (*Cre-LoxP* mediated excision) to make a pituitary-specific knockout of *Gata2*. The mutants are in clear contrast to wild-type mice, and demonstrated fewer thyrotropic cells post natally and reduced function of both gonadotropes and thyrotropes in adult mice [216]. Interestingly, they also discovered that *Gata3* transcript level is elevated in *Gata2* deficient mice, suggesting that GATA2 may negatively regulate the expression of GATA3 [216].

While genetic knockout models provide valuable *in vivo* information, the impact of a *GATA2* threshold in the adult HSC compartment remained unanswered. To address this question, forced expression of GATA2 was conducted by several groups. Indeed, multiple lines of evidence have demonstrated that GATA2 expression levels are a determinant of either self-renewal or differentiation in haematopoietic progenitor cells. For instance, exogenous overexpression of *Gata2* by 2-fold in BM cells can reduce haematopoietic reconstitution at least 40-fold after engraftment [217]. In addition, these mice exhibit cytopaenia and splenomegaly following transplantation of *Gata2* overexpressing BM cells [217]. Another line of evidence was demonstrated by Heyworth *et al*. who recognized that enforced *Gata2* expression through an estrogen inducible GATA2/ER fusion could cease proliferation and trigger differentiation in mouse haematopoietic progenitor cell lines and primary cells [218]. Tipping *et al*., on the other hand, demonstrated that enforced GATA2/ER expression in human cord blood blocks haematopoietic reconstitution after engraftment [219]. The usage of a chimeric TF is arguable and possibly does not fully reflect the actual biological relevance. To address this issue, Kitajima *et al*. used a conditional expression strategy and they showed that GATA2 promote production of haematopoietic progenitor cells. Conversely, GATA2/ER exhibits essentially opposite effects which include inhibition of progenitor colony formation and enhanced differentiation of erythrocyte and megakaryocytes. The authors speculate that the contradictory effect is due to differences in transcriptional interactions and binding activities of the proteins [220]. Nevertheless, these findings highlight a GATA2 dose-dependent regulation of blood cell formation.

GATA family members recognize a specific core binding sequence, WGATAR. GATA2 and GATA3, however, exhibit a broader DNA binding property compared to other GATA members. Besides binding to WGATAR consensus sequence, these factors also recognize an alternative DNA sequence motif consisting of AGATCTTA [221]. Given the distinct, but overlapping, expression patterns of GATA family members, an alternative DNA binding element may help to distinguish GATA2 and GATA3 from other family members in their unique regulatory roles *in vivo* [222]. Collectively, these results mirror the broader role of GATA2 and GATA3 than was previously envisioned.

GATA2 and GATA3 play essential roles in the development of inner ear [204, 223]. *Gata2* is expressed predominantly in the dorsal ventricular system, while *Gata3* is expressed primarily in the ventral cochlear duct and ganglion [223]. Surprisingly, no defects were detected in the inner ear of *Gata2*$^{-/-}$ mouse embryos which died before E10.5. In contrast, *Gata3* null embryos exhibited profound defects in ear development [224]. To circumvent the embryonic lethality phenotype, Finnish scientists generated a *Cre-LoxP* mediated conditional *Gata2* allele knockout lines [225]. These *Gata2*$^{-/-}$ mouse embryos had malformed semicircular ducts and were unable to generate proper perilymphatic space. This was the first evidence demonstrating GATA2 as a critical factor for the clearing of the vestibular perilymphatic mesenchyme [225]. In summary, extensive evidence shows that GATA2 is a key regulator governing many mammalian developmental processes.

1.3.1.3 GATA3

GATA3 was initially isolated from a chicken cDNA library [199]. *GATA3* is the main GATA family member in the mammalian immune system in regulating T-cell differentiation and maturation, particularly in developing and mature T cells, natural killer (NK) cells, CD1-restricted natural killer T cells (NKT) [226-228] and T helper 2 (Th2) cells [229]. Several studies have reported expression of GATA3 in definitive erythrocytes and T lymphocytes, and many non-haematopoietic tissues [199, 230-233], suggesting its specific roles in development and organogenesis. Collectively, these studies provide important contribution to our understanding of GATA3 in T cell developments.

1.3.2 GATA1 and 2 Related Diseases

Perturbation of transcription factors regulating proliferation, differentiation, survival and apoptosis is linked to malignant transformation and other diseases. Given the indispensable roles of GATA family members in normal cellular functions and developmental processes, it is not surprising that disruption of these transcription factors is associated with many diseases. Nichols and co-workers discovered a *GATA1* missense mutation in a family affected with X-linked dyserythropoietic anaemia and thrombocytopenia in 2000 [234]. Later studies have extended and confirmed that other germline missense mutations located in the GATA1 N-terminal also manifest abnormal haematopoiesis [235-237]. 2 years following Nichols'

discovery, Wechsler *et al*. described somatic mutations resulting in the exclusive expression of a truncated GATA1 protein in Down syndrome–related acute megakaryoblastic leukemia (DS-AMKL) patients [238]. Approximately 10% of DS infants develop a transient myeloproliferative syndrome (TMS) [239]. However, nearly one-fifth of them eventually evolve into DS-AMKL later in their life [240, 241]. Triggered by the close connection between DS-AMKL and DS related TMS (DS-TMS), Greene and co-workers investigated genetic alterations in the GATA1 of the DS-TMS patients [242]. They detected marked GATA1 lesions in every DS-TMS infant, contributing to the hypothesis that GATA1 mutation is an initial and important step in DS-AMKL.

Given that GATA2 is a critical regulator in haematopoiesis, any disruption and deregulation in the context of aberrant expression or mutation may be catastrophic. The first evidence linking *GATA2* to MDS and AML was through the analysis of 3q21 chromosome breakpoints [243] which resulted in enhanced GATA2 expression in these patients [243]. Subsequent studies have also shown that GATA2 overexpression is common in AML and MDS patients regardless of the presence of the 3q21 rearrangement [244-246]. The question of how *GATA*2 contributes to haematological malignancy was poorly studied until Zhang *et al*. sparked a new interest in the scientific community. They discovered 2 somatic mutations (an in frame deletion of 6 amino acids and a L359V missense mutation) in CML-BC patients [138]. Both mutations impair myelomonocytic differentiation and perturb transactivation of target genes and DNA binding. To date, these mutations are restricted in CML-BC and have been not found in AML, MDS, ALL or CLL patients [138]. Since then, GATA2 has been actively studied by many researchers. In 2011 and 2012, there were more than 10 publications regarding novel *GATA2* mutations. Most of these mutations are LOF mutants and are associated with a higher tendency to develop AML-M5 [247], MDS and MDS/AML [153], Emberger syndrome [162], MonoMac syndrome [166] and DCML deficiency [167]. We first described 2 novel germline GATA2 mutations (T354M and T355del) in 4 separate families affected with familial MDS or MDS/AML [153, 165] (see **Chapter 3**). Outside of the haematopoietic and lymphatic vasculature system, GATA2 has been implicated in atherosclerosis [248], human breast cancer [249], prostate cancer [250], early-onset coronary

artery disease [251] and Parkinson's disease [252], although very little is known about the mechanism of its action.

1.4 Aim of the Study

Leukaemia is a disease that causes profound impact on economic and social burden worldwide. Despite advances in modern medicine, treatment of the disease, particularly in elderly AML patients, remains difficult with significant therapy-related morbidity. Currently, the front line treatment for AML is chemotherapy. Other treatment options include allogeneic and autologous BM transplantation, and radiotherapy.

Leukaemia is thought to manifest as the consequence of combinatorial interactions between inherited and acquired genetic factors and various environmental factors throughout a lifetime. At the commencement of my study (2009), there were two well described familial MDS/AML predisposing genes, *RUNX1* and *CEBPA*, and the newly identified *TERT* and *TERC*. However, there are still a number of families negative for these mutations. In the quest for searching for a novel predisposition genes in familial AML, a candidate gene targeted sequencing strategy will be employed on several families affected with hereditary myeloid malignancies. With special focus on the myeloid compartment, we will then seek to understand the molecular mechanism driven by the mutated gene(s) through a series of *in vitro* and *ex vivo* functional assays. In summary, this study aims to identify a novel predisposition gene for familial MDS/AML and to decipher the molecular mechanism underlying disease development.

1.5 References

1. Rodak, B.F., G.A. Fritsma, and K. Doig, *Hematology: Clinical Principles and Applications*. 2007: Saunders Elsevier. p.86.
2. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology.* Cell, 2008. **132**(4): p. 631-44.
3. Kondo, M., et al., *Biology of hematopoietic stem cells and progenitors: implications for clinical application.* Annu Rev Immunol, 2003. **21**: p. 759-806.
4. Proytcheva, M.A., *Diagnostic Pediatric Hematopathology*. 2011: Cambridge University Press.
5. Galloway, J.L. and L.I. Zon, *Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo.* Curr Top Dev Biol, 2003. **53**: p. 139-58.
6. Cumano, A., et al., *Intraembryonic, but not yolk sac hematopoietic precursors, isolated before circulation, provide long-term multilineage reconstitution.* Immunity, 2001. **15**(3): p. 477-85.
7. Medvinsky, A. and E. Dzierzak, *Definitive hematopoiesis is autonomously initiated by the AGM region.* Cell, 1996. **86**(6): p. 897-906.
8. de Bruijn, M.F., et al., *Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo.* The EMBO journal, 2000. **19**(11): p. 2465-74.
9. Gekas, C., et al., *The placenta is a niche for hematopoietic stem cells.* Dev Cell, 2005. **8**(3): p. 365-75.
10. Dame, C. and S.E. Juul, *The switch from fetal to adult erythropoiesis.* Clin Perinatol, 2000. **27**(3): p. 507-26.
11. Zhang, J. and L. Li, *Stem cell niche: microenvironment and beyond.* J Biol Chem, 2008. **283**(15): p. 9499-503.
12. Morrison, S.J. and I.L. Weissman, *The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype.* Immunity, 1994. **1**(8): p. 661-73.
13. Wilson, A. and A. Trumpp, *Bone-marrow haematopoietic-stem-cell niches.* Nat Rev Immunol, 2006. **6**(2): p. 93-106.
14. Reya, T., et al., *Stem cells, cancer, and cancer stem cells.* Nature, 2001. **414**(6859): p. 105-11.
15. Yang, L., et al., *Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients.* Blood, 2005. **105**(7): p. 2717-23.
16. Christensen, J.L. and I.L. Weissman, *Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14541-6.
17. Morrison, S.J., et al., *Identification of a lineage of multipotent hematopoietic progenitors.* Development, 1997. **124**(10): p. 1929-39.
18. Kondo, M., I.L. Weissman, and K. Akashi, *Identification of clonogenic common lymphoid progenitors in mouse bone marrow.* Cell, 1997. **91**(5): p. 661-72.
19. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-7.

20. Adolfsson, J., et al., *Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity.* Immunity, 2001. **15**(4): p. 659-69.
21. Adolfsson, J., et al., *Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment.* Cell, 2005. **121**(2): p. 295-306.
22. Lai, A.Y. and M. Kondo, *Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors.* The Journal of experimental medicine, 2006. **203**(8): p. 1867-73.
23. Lai, A.Y. and M. Kondo, *T and B lymphocyte differentiation from hematopoietic stem cell.* Semin Immunol, 2008. **20**(4): p. 207-12.
24. Handin, R.I., S.E. Lux, and T.P. Stossel, *Blood: Principles and Practice of Hematology*. 2003: Lippincott Williams & Wilkins. p.189.
25. Sieweke, M.H. and T. Graf, *A transcription factor party during blood cell differentiation.* Curr Opin Genet Dev, 1998. **8**(5): p. 545-51.
26. Begley, C.G., et al., *The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif.* Proc Natl Acad Sci U S A, 1989. **86**(24): p. 10128-32.
27. Aggarwal, R., et al., *Hematopoietic stem cells: transcriptional regulation, ex vivo expansion and clinical application.* Curr Mol Med, 2012. **12**(1): p. 34-49.
28. Hall, M.A., et al., *The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 992-7.
29. Robb, L., et al., *Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 7075-9.
30. Okuda, T., et al., *AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis.* Cell, 1996. **84**(2): p. 321-30.
31. Tsai, S.F., et al., *Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells.* Nature, 1989. **339**(6224): p. 446-51.
32. Mikkola, H.K., et al., *Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene.* Nature, 2003. **421**(6922): p. 547-51.
33. Visvader, J.E., Y. Fujiwara, and S.H. Orkin, *Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development.* Genes Dev, 1998. **12**(4): p. 473-9.
34. North, T., et al., *Cbfa2 is required for the formation of intra-aortic hematopoietic clusters.* Development, 1999. **126**(11): p. 2563-75.
35. North, T.E., et al., *Runx1 is expressed in adult mouse hematopoietic stem cells and differentiating myeloid and lymphoid cells, but not in maturing erythroid cells.* Stem Cells, 2004. **22**(2): p. 158-68.
36. Wang, Q., et al., *Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis.* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3444-9.
37. Growney, J.D., et al., *Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype.* Blood, 2005. **106**(2): p. 494-504.
38. Ichikawa, M., et al., *AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis.* Nat Med, 2004. **10**(3): p. 299-304.

39. Orlic, D., et al., *Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45 NF-E2, and c-myb and low levels or no mRNA for c-fms and the receptors for granulocyte colony-stimulating factor and interleukins 5 and 7.* Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4601-5.
40. Wadman, I.A., et al., *The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins.* The EMBO journal, 1997. **16**(11): p. 3145-57.
41. Lecuyer, E., et al., *The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1.* Blood, 2002. **100**(7): p. 2430-40.
42. Pimanda, J.E., et al., *Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development.* Proc Natl Acad Sci U S A, 2007. **104**(45): p. 17692-7.
43. Wilson, N.K., et al., *Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.* Cell Stem Cell, 2010. **7**(4): p. 532-44.
44. Love, P., *Genes and Signals Regulating Mammalian Hematopoiesis, Annual Report of the Division of Intramural Research, Available at http://2010annualreport.nichd.nih.gov/scdb.html*, 2010, NIH.
45. Klemsz, M.J., et al., *The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene.* Cell, 1990. **61**(1): p. 113-24.
46. Nutt, S.L., et al., *Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors.* The Journal of experimental medicine, 2005. **201**(2): p. 221-31.
47. Rosenbauer, F., et al., *Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1.* Nature genetics, 2006. **38**(1): p. 27-37.
48. Dahl, R., et al., *Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor.* Nat Immunol, 2003. **4**(10): p. 1029-36.
49. Hohaus, S., et al., *PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene.* Mol Cell Biol, 1995. **15**(10): p. 5830-45.
50. Pahl, H.L., et al., *The proto-oncogene PU.1 regulates expression of the myeloid-specific CD11b promoter.* J Biol Chem, 1993. **268**(7): p. 5014-20.
51. Zhang, D.E., et al., *The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor.* Mol Cell Biol, 1994. **14**(1): p. 373-81.
52. Cook, W.D., et al., *PU.1 is a suppressor of myeloid leukemia, inactivated in mice by gene deletion and mutation of its DNA binding domain.* Blood, 2004. **104**(12): p. 3437-44.
53. Mueller, B.U., et al., *Heterozygous PU.1 mutations are associated with acute myeloid leukemia.* Blood, 2002. **100**(3): p. 998-1007.
54. Radomska, H.S., et al., *CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors.* Mol Cell Biol, 1998. **18**(7): p. 4301-14.
55. Scott, L.M., et al., *A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells.* Blood, 1992. **80**(7): p. 1725-35.

56. Lin, F.T., et al., *A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity.* Proc Natl Acad Sci U S A, 1993. **90**(20): p. 9606-10.
57. Nerlov, C., *C/EBPalpha mutations in acute myeloid leukaemias.* Nat Rev Cancer, 2004. **4**(5): p. 394-400.
58. Zhang, D.E., et al., *Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice.* Proc Natl Acad Sci U S A, 1997. **94**(2): p. 569-74.
59. Morosetti, R., et al., *A novel, myeloid transcription factor, C/EBP epsilon, is upregulated during granulocytic, but not monocytic, differentiation.* Blood, 1997. **90**(7): p. 2591-600.
60. Zhang, P., et al., *Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha.* Immunity, 2004. **21**(6): p. 853-63.
61. Mueller, B.U. and T. Pabst, *C/EBPalpha and the pathophysiology of acute myeloid leukemia.* Curr Opin Hematol, 2006. **13**(1): p. 7-14.
62. Seth, A., et al., *Transactivation of GATA-1 promoter with ETS1, ETS2 and ERGB/Hu-FLI-1 proteins: stabilization of the ETS1 protein binding on GATA-1 promoter sequences by monoclonal antibody.* Oncogene, 1993. **8**(7): p. 1783-90.
63. Watson, D.K., et al., *The ERGB/Fli-1 gene: isolation and characterization of a new member of the family of human ETS transcription factors.* Cell Growth Differ, 1992. **3**(10): p. 705-13.
64. Hart, A., et al., *Fli-1 is required for murine vascular and megakaryocytic development and is hemizygously deleted in patients with thrombocytopenia.* Immunity, 2000. **13**(2): p. 167-77.
65. Spyropoulos, D.D., et al., *Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor.* Mol Cell Biol, 2000. **20**(15): p. 5643-52.
66. Starck, J., et al., *Inducible Fli-1 gene deletion in adult mice modifies several myeloid lineage commitment decisions and accelerates proliferation arrest and terminal erythrocytic differentiation.* Blood, 2010. **116**(23): p. 4795-805.
67. Masuda, A., et al., *Essential role of GATA transcriptional factors in the activation of mast cells.* Journal of immunology, 2007. **178**(1): p. 360-8.
68. Zhang, X.K., et al., *The transcription factor Fli-1 modulates marginal zone and follicular B cell development in mice.* Journal of immunology, 2008. **181**(3): p. 1644-54.
69. AIHW, *Cancer in Australia: an overview, 2008 , Cancer series no. 46. Cat. no. CAN 42. Available at http://www.aihw.gov.au/publications/index.cfm/title/10607*. 2008, Canberra: Australian Institute of Health and Welfare.
70. AIHW, *Cancer survival and prevalence in Australia : Cancers diagnosed from 1982 to 2004, Cat. no. CAN 38. Available at http://www.aihw.gov.au/publications/index.cfm/title/10484*. 2008, Canberra: The Australian Institute of Health and Welfare. 88.
71. Howlader N, N.A., Krapcho M, Neyman N, Aminou R, Waldron W, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, Cronin KA, Edwards BK (eds). , *SEER Cancer Statistics Review, 1975-2008, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2008/,*

based on November 2010 SEER data submission, posted to the SEER web site, 2011., 2011.
72. Bennett, J.M., et al., *Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group.* British journal of haematology, 1976. **33**(4): p. 451-8.
73. Bennett, J.M., et al., *Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group.* Annals of internal medicine, 1985. **103**(4): p. 620-5.
74. Haferlach, T., et al., *Morphologic dysplasia in de novo acute myeloid leukemia (AML) is related to unfavorable cytogenetics but has no independent prognostic relevance under the conditions of intensive induction therapy: results of a multiparameter analysis from the German AML Cooperative Group studies.* J Clin Oncol, 2003. **21**(2): p. 256-65.
75. Elaine S. Jaffe, et al., *Classification of Tumours: Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues.* 2001, Lyon, France: IARC Press.
76. Vardiman, J.W., N.L. Harris, and R.D. Brunning, *The World Health Organization (WHO) classification of the myeloid neoplasms.* Blood, 2002. **100**(7): p. 2292-302.
77. Vardiman, J.W., *The World Health Organization (WHO) classification of tumors of the hematopoietic and lymphoid tissues: an overview with emphasis on the myeloid neoplasms.* Chemico-biological interactions, 2010. **184**(1-2): p. 16-20.
78. Koschmieder, S., et al., *Dysregulation of the C/EBPalpha differentiation pathway in human cancer.* J Clin Oncol, 2009. **27**(4): p. 619-28.
79. Mrozek, K. and C.D. Bloomfield, *Chromosome aberrations, gene mutations and expression changes, and prognosis in adult acute myeloid leukemia.* Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program, 2006: p. 169-77.
80. Schotte, D., R. Pieters, and M.L. Den Boer, *MicroRNAs in acute leukemia: from biological players to clinical contributors.* Leukemia, 2012. **26**(1): p. 1-12.
81. Ghanem, H., N. Tank, and I.A. Tabbara, *Prognostic implications of genetic aberrations in acute myelogenous leukemia with normal cytogenetics.* Am J Hematol, 2012. **87**(1): p. 69-77.
82. Huret, J.L., et al., *Atlas of genetics and cytogenetics in oncology and haematology, an interactive database.* Nucleic acids research, 2000. **28**(1): p. 349-51.
83. Rossetti, S. and N. Sacchi, *RUNX1: A MicroRNA Hub in Normal and Malignant Hematopoiesis.* Int J Mol Sci, 2013. **14**(1): p. 1566-88.
84. Roumier, C., et al., *New mechanisms of AML1 gene alteration in hematological malignancies.* Leukemia, 2003. **17**(1): p. 9-16.
85. Osato, M., *Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia.* Oncogene, 2004. **23**(24): p. 4284-96.
86. Song, W.J., et al., *Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia.* Nature genetics, 1999. **23**(2): p. 166-75.
87. Harada, H., et al., *High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia.* Blood, 2004. **103**(6): p. 2316-24.
88. Leroy, H., et al., *CEBPA point mutations in hematological malignancies.* Leukemia, 2005. **19**(3): p. 329-34.

89. Schlenk, R.F., et al., *Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia.* The New England journal of medicine, 2008. **358**(18): p. 1909-18.
90. Sellick, G.S., et al., *Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia.* Leukemia, 2005. **19**(7): p. 1276-8.
91. Smith, M.L., et al., *Mutation of CEBPA in familial acute myeloid leukemia.* The New England journal of medicine, 2004. **351**(23): p. 2403-7.
92. Pabst, T., et al., *Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia.* Nature genetics, 2001. **27**(3): p. 263-70.
93. Asou, H., et al., *Establishment of the acute myeloid leukemia cell line Kasumi-6 from a patient with a dominant-negative mutation in the DNA-binding region of the C/EBPalpha gene.* Genes Chromosomes Cancer, 2003. **36**(2): p. 167-74.
94. Gombart, A.F., et al., *Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias.* Blood, 2002. **99**(4): p. 1332-40.
95. Dufour, A., et al., *Monoallelic CEBPA mutations in normal karyotype acute myeloid leukemia: independent favorable prognostic factor within NPM1 mutated patients.* Annals of hematology, 2012. **91**(7): p. 1051-63.
96. Owen, C., M. Barnett, and J. Fitzgibbon, *Familial myelodysplasia and acute myeloid leukaemia--a review.* Br J Haematol, 2008. **140**(2): p. 123-32.
97. Bereshchenko, O., et al., *Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML.* Cancer Cell, 2009. **16**(5): p. 390-400.
98. Kato, N., et al., *Two types of C/EBPalpha mutations play distinct but collaborative roles in leukemogenesis: lessons from clinical data and BMT models.* Blood, 2011. **117**(1): p. 221-33.
99. Marchesi, F., et al., *Cytogenetic abnormalities in adult non-promyelocytic acute myeloid leukemia: a concise review.* Crit Rev Oncol Hematol, 2011. **80**(3): p. 331-46.
100. Reikvam, H., et al., *Acute myeloid leukemia with the t(8;21) translocation: clinical consequences and biological implications.* J Biomed Biotechnol, 2011. **2011**: p. 104631.
101. Le Beau, M.M., et al., *Association of an inversion of chromosome 16 with abnormal marrow eosinophils in acute myelomonocytic leukemia. A unique cytogenetic-clinicopathological association.* The New England journal of medicine, 1983. **309**(11): p. 630-6.
102. Lowenberg, B., J.R. Downing, and A. Burnett, *Acute myeloid leukemia.* The New England journal of medicine, 1999. **341**(14): p. 1051-62.
103. Rowley, J.D., *The critical role of chromosome translocations in human leukemias.* Annu Rev Genet, 1998. **32**: p. 495-519.
104. Jones, D., *Neoplastic hematopathology : experimental and clinical approaches*. Contemporary hematology. 2010, New York: Humana Press. p156.
105. James R. Downing, F.G.B., *Neoplastic Hematopathology*. 2nd ed, ed. D.M. Knowles. 2001, Philadephia: Lippincott Williams & Wilkins. p.1626.
106. Moqattash, S. and J.D. Lutton, *Leukemia cells and the cytokine network.* Proc Soc Exp Biol Med, 1998. **219**(1): p. 8-27.

107. Christiansen, D.H., et al., *Mutations of genes in the receptor tyrosine kinase (RTK)/RAS-BRAF signal transduction pathway in therapy-related myelodysplasia and acute myeloid leukemia.* Leukemia, 2005. **19**(12): p. 2232-40.
108. Pedersen-Bjergaard, J., M.T. Andersen, and M.K. Andersen, *Genetic pathways in the pathogenesis of therapy-related myelodysplasia and acute myeloid leukemia.* Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program, 2007: p. 392-7.
109. Mauritzson, N., et al., *Pooled analysis of clinical and cytogenetic features in treatment-related and de novo adult acute myeloid leukemia and myelodysplastic syndromes based on a consecutive series of 761 patients analyzed 1976-1993 and on 5098 unselected cases reported in the literature 1974-2001.* Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 2002. **16**(12): p. 2366-78.
110. Pedersen-Bjergaard, J., et al., *Different genetic pathways in leukemogenesis for patients presenting with therapy-related myelodysplasia and therapy-related acute myeloid leukemia.* Blood, 1995. **86**(9): p. 3542-52.
111. Arthur, D.C., et al., *The clinical significance of karyotype in acute myelogenous leukemia.* Cancer genetics and cytogenetics, 1989. **40**(2): p. 203-16.
112. Estey, E.H., S. Pierce, and M.J. Keating, *Identification of a group of AML/MDS patients with a relatively favorable prognosis who have chromosome 5 and/or 7 abnormalities.* Haematologica, 2000. **85**(3): p. 246-9.
113. Swansbury, G.J., et al., *Long-term survival in acute myelogenous leukemia: a second follow-up of the Fourth International Workshop on Chromosomes in Leukemia.* Cancer genetics and cytogenetics, 1994. **73**(1): p. 1-7.
114. Srivastava, S., D.E. Henson, and A.F. Gazdar, *Molecular pathology of early cancer*. 1999, Amsterdam ; Washington, D.C.: IOS Press. xix, 481 p.
115. Rowe, J.M., *Optimal induction and post-remission therapy for AML in first remission.* Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program, 2009: p. 396-405.
116. Byrd, J.C., et al., *Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461).* Blood, 2002. **100**(13): p. 4325-36.
117. Grimwade, D., et al., *The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties.* Blood, 1998. **92**(7): p. 2322-33.
118. Pedersen-Bjergaard, J. and J.D. Rowley, *The balanced and the unbalanced chromosome aberrations of acute myeloid leukemia may develop in different ways and may contribute differently to malignant transformation.* Blood, 1994. **83**(10): p. 2780-6.
119. Dastugue, N., et al., *Prognostic significance of karyotype in de novo adult acute myeloid leukemia. The BGMT group.* Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 1995. **9**(9): p. 1491-8.
120. Leith, C.P., et al., *Acute myeloid leukemia in the elderly: assessment of multidrug resistance (MDR1) and cytogenetics distinguishes biologic subgroups with remarkably distinct responses to standard chemotherapy. A Southwest Oncology Group study.* Blood, 1997. **89**(9): p. 3323-9.

121. Corey, S.J., et al., *Myelodysplastic syndromes: the complexity of stem-cell diseases.* Nature reviews. Cancer, 2007. **7**(2): p. 118-29.
122. Armitage, J.O., *Atlas of clinical hematology*. 2003, Philadelphia: Current Medicine. p.266.
123. Siegenthaler, W., *Differential Diagnosis in Internal Medicine: From Symptom to Diagnosis*. 2007: Thieme. p.423-433.
124. Germing, U., et al., *No increase in age-specific incidence of myelodysplastic syndromes.* Haematologica, 2004. **89**(8): p. 905-10.
125. Parker, J.E. and G.J. Mufti, *The role of apoptosis in the pathogenesis of the myelodysplastic syndromes.* International journal of hematology, 2001. **73**(4): p. 416-28.
126. Liew, E. and C. Owen, *Familial myelodysplastic syndromes: a review of the literature.* Haematologica, 2011. **96**(10): p. 1536-42.
127. Padua, R.A., A. McGlynn, and H. McGlynn, *Molecular, cytogenetic and genetic abnormalities in MDS and secondary AML.* Cancer Treat Res, 2001. **108**: p. 111-57.
128. Kundranda, M.N., R. Tibes, and R.A. Mesa, *Transformation of a chronic myeloproliferative neoplasm to acute myelogenous leukemia: does anything work?* Curr Hematol Malig Rep, 2012. **7**(1): p. 78-86.
129. Barbui, T. and A. Tefferi, *Myeloproliferative Neoplasms*. 2012: Springer Berlin Heidelberg. p.4.
130. Dameshek, W., *Some speculations on the myeloproliferative syndromes.* Blood, 1951. **6**(4): p. 372-5.
131. Rowley, J.D., *Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining.* Nature, 1973. **243**(5405): p. 290-3.
132. Shtivelman, E., et al., *Fused transcript of abl and bcr genes in chronic myelogenous leukaemia.* Nature, 1985. **315**(6020): p. 550-4.
133. Davis, R.L., J.B. Konopka, and O.N. Witte, *Activation of the c-abl oncogene by viral transduction or chromosomal translocation generates altered c-abl proteins with similar in vitro kinase properties.* Mol Cell Biol, 1985. **5**(1): p. 204-13.
134. Kelliher, M.A., et al., *Induction of a chronic myelogenous leukemia-like syndrome in mice with v-abl and BCR/ABL.* Proc Natl Acad Sci U S A, 1990. **87**(17): p. 6649-53.
135. Feinstein, E., et al., *p53 in chronic myelogenous leukemia in acute phase.* Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6293-7.
136. Sill, H., J.M. Goldman, and N.C. Cross, *Homozygous deletions of the p16 tumor-suppressor gene are associated with lymphoid transformation of chronic myeloid leukemia.* Blood, 1995. **85**(8): p. 2013-6.
137. Towatari, M., et al., *Absence of the human retinoblastoma gene product in the megakaryoblastic crisis of chronic myelogenous leukemia.* Blood, 1991. **78**(9): p. 2178-81.
138. Zhang, S.J., et al., *Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia.* Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2076-81.
139. Wolanskyj, A.P., et al., *Essential thrombocythemia beyond the first decade: life expectancy, long-term complication rates, and prognostic factors.* Mayo Clin Proc, 2006. **81**(2): p. 159-66.

140. Campbell, P.J. and A.R. Green, *The myeloproliferative disorders.* The New England journal of medicine, 2006. **355**(23): p. 2452-66.
141. Finazzi, G., et al., *Acute leukemia in polycythemia vera: an analysis of 1638 patients enrolled in a prospective observational study.* Blood, 2005. **105**(7): p. 2664-70.
142. Dupriez, B., et al., *Prognostic factors in agnogenic myeloid metaplasia: a report on 195 cases with a new scoring system.* Blood, 1996. **88**(3): p. 1013-8.
143. Parker, J.E., et al., *'Low-risk' myelodysplastic syndrome is associated with excessive apoptosis and an increased ratio of pro- versus anti-apoptotic bcl-2-related proteins.* Br J Haematol, 1998. **103**(4): p. 1075-82.
144. Szpurka, H., et al., *Refractory anemia with ringed sideroblasts associated with marked thrombocytosis (RARS-T), another myeloproliferative condition characterized by JAK2 V617F mutation.* Blood, 2006. **108**(7): p. 2173-81.
145. Kouides, P.A. and J.M. Bennett *Morphology and classification of the MDS and their pathologic variants*. 1996. 26.
146. Horwitz, M., *The genetics of familial leukemia.* Leukemia, 1997. **11**(8): p. 1347-59.
147. Benson, K.F. and M. Horwitz, *Familial leukemia.* Best Pract Res Clin Haematol, 2006. **19**(2): p. 269-79.
148. Perry, C., A. Eldor, and H. Soreq, *Runx1/AML1 in leukemia: disrupted association with diverse protein partners.* Leuk Res, 2002. **26**(3): p. 221-8.
149. Dowton, S.B., et al., *Studies of a familial platelet disorder.* Blood, 1985. **65**(3): p. 557-63.
150. Ho, C.Y., et al., *Linkage of a familial platelet disorder with a propensity to develop myeloid malignancies to human chromosome 21q22.1-22.2.* Blood, 1996. **87**(12): p. 5218-24.
151. Nanri, T., et al., *A family harboring a germ-line N-terminal C/EBPalpha mutation and development of acute myeloid leukemia with an additional somatic C-terminal C/EBPalpha mutation.* Genes Chromosomes Cancer, 2010. **49**(3): p. 237-41.
152. Carmichael, C.L., et al., *Poor prognosis in familial acute myeloid leukaemia with combined biallelic CEBPA mutations and downstream events affecting the ATM, FLT3 and CDX2 genes.* Br J Haematol, 2010. **150**(3): p. 382-5.
153. Hahn, C.N., et al., *Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia.* Nature genetics, 2011. **43**(10): p. 1012-7.
154. Carmichael, C.L. and H.S. Scott, *Familial aspects of haematological malignancy.* Cancer Forum, Nov 2007. **31**(3): p. 160-163.
155. Sankaran, V.G., et al., *Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia.* J Clin Invest, 2012. **122**(7): p. 2439-43.
156. Yamaguchi, H., et al., *Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome.* Blood, 2003. **102**(3): p. 916-8.
157. Emberger, J., et al., *Surdimutite, lymphoedema des membres inferieurs et anomalies hematologiques (leucoseaigue, cytopenies) a transmission autosomique dominante.* J GenetHum, 1979. **67**: p. 1382–1388.
158. Attal, M., et al., *Association lymphoedema idiopathique et leucemie aigue familiale.* Presse Med, 1985. **14**: p. 600.
159. Kaur, J., et al., *Familial acute myeloid leukaemia with acquired Pelger-Huet anomaly and aneuploidy of C group.* Br Med J, 1972. **4**(5836): p. 327-31.

160. Mansour, S., et al., *Emberger syndrome-primary lymphedema with myelodysplasia: report of seven new cases.* Am J Med Genet A, 2010. **152A**(9): p. 2287-96.
161. Kazenwadel, J., et al., *Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature.* Blood, 2012. **119**(5): p. 1283-91.
162. Ostergaard, P., et al., *Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome).* Nature genetics, 2011. **43**(10): p. 929-31.
163. Bigley, V., et al., *The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency.* The Journal of experimental medicine, 2011. **208**(2): p. 227-34.
164. Vinh, D.C., et al., *Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia.* Blood, 2010. **115**(8): p. 1519-29.
165. Scott, H.S., et al. *GATA2 is a new predisposition gene for familial myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML) [Late-breaking Abstracts]*. in *American Society of Hematology Annual Meeting*. 2010. Orlando Florida.
166. Hsu, A.P., et al., *Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome.* Blood, 2011. **118**(10): p. 2653-5.
167. Dickinson, R.E., et al., *Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.* Blood, 2011. **118**(10): p. 2656-8.
168. Zon, L.I., et al., *Expression of GATA-binding proteins during embryonic development in Xenopus laevis.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10642-6.
169. Weiss, M.J. and S.H. Orkin, *GATA transcription factors: key regulators of hematopoiesis.* Exp Hematol, 1995. **23**(2): p. 99-107.
170. Shimizu, R., J.D. Engel, and M. Yamamoto, *GATA1-related leukaemias.* Nat Rev Cancer, 2008. **8**(4): p. 279-87.
171. Sigvardsson, M., *New light on the biology and developmental potential of haematopoietic stem cells and progenitor cells.* J Intern Med, 2009. **266**(4): p. 311-24.
172. Ting, C.N., et al., *Transcription factor GATA-3 is required for development of the T-cell lineage.* Nature, 1996. **384**(6608): p. 474-8.
173. Molkentin, J.D., *The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression.* J Biol Chem, 2000. **275**(50): p. 38949-52.
174. Zon, L.I., et al., *The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome.* Proc Natl Acad Sci U S A, 1990. **87**(2): p. 668-72.
175. Evans, T. and G. Felsenfeld, *The erythroid-specific transcription factor Eryf1: a new finger protein.* Cell, 1989. **58**(5): p. 877-85.
176. Evans, T., M. Reitman, and G. Felsenfeld, *An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes.* Proc Natl Acad Sci U S A, 1988. **85**(16): p. 5976-80.
177. Fujiwara, Y., et al., *Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12355-8.

178. Leonard, M., et al., *Dynamics of GATA transcription factor expression during erythroid differentiation.* Blood, 1993. **82**(4): p. 1071-9.
179. Martin, D.I., et al., *Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages.* Nature, 1990. **344**(6265): p. 444-7.
180. Zon, L.I., et al., *Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription.* Blood, 1993. **81**(12): p. 3234-41.
181. Ito, E., et al., *Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis.* Nature, 1993. **362**(6419): p. 466-8.
182. Yomogida, K., et al., *Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells.* Development, 1994. **120**(7): p. 1759-66.
183. Pevny, L., et al., *Development of hematopoietic cells lacking transcription factor GATA-1.* Development, 1995. **121**(1): p. 163-72.
184. Pevny, L., et al., *Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1.* Nature, 1991. **349**(6306): p. 257-60.
185. Weiss, M.J., G. Keller, and S.H. Orkin, *Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells.* Genes & development, 1994. **8**(10): p. 1184-97.
186. Weiss, M.J. and S.H. Orkin, *Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis.* Proc Natl Acad Sci U S A, 1995. **92**(21): p. 9623-7.
187. Takahashi, S., et al., *Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene.* The Journal of biological chemistry, 1997. **272**(19): p. 12611-5.
188. Takahashi, S., et al., *Role of GATA-1 in proliferation and differentiation of definitive erythroid and megakaryocytic cells in vivo.* Blood, 1998. **92**(2): p. 434-42.
189. Takahashi, S., et al., *GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies.* Blood, 2000. **96**(3): p. 910-6.
190. Visvader, J.E., et al., *GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line.* The EMBO journal, 1992. **11**(12): p. 4557-64.
191. Kulessa, H., J. Frampton, and T. Graf, *GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts.* Genes & development, 1995. **9**(10): p. 1250-62.
192. Vyas, P., et al., *Consequences of GATA-1 deficiency in megakaryocytes and platelets.* Blood, 1999. **93**(9): p. 2867-75.
193. Hirasawa, R., et al., *Essential and instructive roles of GATA factors in eosinophil development.* The Journal of experimental medicine, 2002. **195**(11): p. 1379-86.
194. Yu, C., et al., *Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo.* The Journal of experimental medicine, 2002. **195**(11): p. 1387-95.
195. Tsai, S.F., E. Strauss, and S.H. Orkin, *Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter.* Genes Dev, 1991. **5**(6): p. 919-31.

196. Zon, L.I., et al., *GATA-binding transcription factors in mast cells regulate the promoter of the mast cell carboxypeptidase A gene.* The Journal of biological chemistry, 1991. **266**(34): p. 22948-53.
197. Harigae, H., et al., *Differential roles of GATA-1 and GATA-2 in growth and differentiation of mast cells.* Genes to cells : devoted to molecular & cellular mechanisms, 1998. **3**(1): p. 39-50.
198. Migliaccio, A.R., et al., *GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant.* The Journal of experimental medicine, 2003. **197**(3): p. 281-96.
199. Yamamoto, M., et al., *Activity and tissue-specific expression of the transcription factor NF-E1 multigene family.* Genes Dev, 1990. **4**(10): p. 1650-62.
200. Visvader, J. and J.M. Adams, *Megakaryocytic differentiation induced in 416B myeloid cells by GATA-2 and GATA-3 transgenes or 5-azacytidine is tightly coupled to GATA-1 expression.* Blood, 1993. **82**(5): p. 1493-501.
201. Laverriere, A.C., et al., *GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut.* J Biol Chem, 1994. **269**(37): p. 23177-84.
202. Minegishi, N., et al., *The mouse GATA-2 gene is expressed in the para-aortic splanchnopleura and aorta-gonads and mesonephros region.* Blood, 1999. **93**(12): p. 4196-207.
203. Perez-Stable, C.M., A. Pozas, and B.A. Roos, *A role for GATA transcription factors in the androgen regulation of the prostate-specific antigen gene enhancer.* Mol Cell Endocrinol, 2000. **167**(1-2): p. 43-53.
204. Lillevali, K., et al., *Comparative analysis of Gata3 and Gata2 expression during chicken inner ear development.* Dev Dyn, 2007. **236**(1): p. 306-13.
205. Kala, K., et al., *Gata2 is a tissue-specific post-mitotic selector gene for midbrain GABAergic neurons.* Development, 2009. **136**(2): p. 253-62.
206. Okitsu, Y., et al., *Regulation of adipocyte differentiation of bone marrow stromal cells by transcription factor GATA-2.* Biochem Biophys Res Commun, 2007. **364**(2): p. 383-7.
207. Wozniak, R.J., et al., *Context-dependent GATA factor function: combinatorial requirements for transcriptional control in hematopoietic and endothelial cells.* J Biol Chem, 2007. **282**(19): p. 14665-74.
208. Minegishi, N., et al., *Alternative promoters regulate transcription of the mouse GATA-2 gene.* The Journal of biological chemistry, 1998. **273**(6): p. 3625-34.
209. Pan, X., et al., *Identification of human GATA-2 gene distal IS exon and its expression in hematopoietic stem cell fractions.* J Biochem, 2000. **127**(1): p. 105-12.
210. Minegishi, N., et al., *Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos.* Blood, 2003. **102**(3): p. 896-905.
211. Suzuki, N., et al., *Combinatorial Gata2 and Sca1 expression defines hematopoietic stem cells in the bone marrow niche.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2202-7.
212. Tsai, F.Y., et al., *An early haematopoietic defect in mice lacking the transcription factor GATA-2.* Nature, 1994. **371**(6494): p. 221-6.
213. Ling, K.W., et al., *GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells.* The Journal of experimental medicine, 2004. **200**(7): p. 871-82.

214. Rodrigues, N.P., et al., *Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis.* Blood, 2005. **106**(2): p. 477-84.
215. Rodrigues, N.P., et al., *GATA-2 regulates granulocyte-macrophage progenitor cell function.* Blood, 2008. **112**(13): p. 4862-73.
216. Charles, M.A., et al., *Pituitary-specific Gata2 knockout: effects on gonadotrope and thyrotrope function.* Mol Endocrinol, 2006. **20**(6): p. 1366-77.
217. Persons, D.A., et al., *Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis.* Blood, 1999. **93**(2): p. 488-99.
218. Heyworth, C., et al., *A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal.* Genes Dev, 1999. **13**(14): p. 1847-60.
219. Tipping, A.J., et al., *High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle.* Blood, 2009. **113**(12): p. 2661-72.
220. Kitajima, K., et al., *GATA-2 and GATA-2/ER display opposing activities in the development and differentiation of blood progenitors.* The EMBO journal, 2002. **21**(12): p. 3060-9.
221. Ko, L.J. and J.D. Engel, *DNA-binding specificities of the GATA transcription factor family.* Mol Cell Biol, 1993. **13**(7): p. 4011-22.
222. Pedone, P.V., et al., *The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains.* The EMBO journal, 1997. **16**(10): p. 2874-82.
223. Lillevali, K., et al., *Partially overlapping expression of Gata2 and Gata3 during inner ear development.* Dev Dyn, 2004. **231**(4): p. 775-81.
224. Karis, A., et al., *Transcription factor GATA-3 alters pathway selection of olivocochlear neurons and affects morphogenesis of the ear.* J Comp Neurol, 2001. **429**(4): p. 615-30.
225. Haugas, M., et al., *Gata2 is required for the development of inner ear semicircular ducts and the surrounding perilymphatic space.* Dev Dyn, 2010. **239**(9): p. 2452-69.
226. Ho, I.C., et al., *Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene.* The EMBO journal, 1991. **10**(5): p. 1187-92.
227. Oosterwegel, M., et al., *Expression of GATA-3 during lymphocyte differentiation and mouse embryogenesis.* Developmental immunology, 1992. **3**(1): p. 1-11.
228. Samson, S.I., et al., *GATA-3 promotes maturation, IFN-gamma production, and liver-specific homing of NK cells.* Immunity, 2003. **19**(5): p. 701-11.
229. Zhang, D.H., et al., *Transcription factor GATA-3 is differentially expressed in murine Th1 and Th2 cells and controls Th2-specific expression of the interleukin-5 gene.* J Biol Chem, 1997. **272**(34): p. 21597-603.
230. George, K.M., et al., *Embryonic expression and cloning of the murine GATA-3 gene.* Development, 1994. **120**(9): p. 2673-86.
231. Pandolfi, P.P., et al., *Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis.* Nature genetics, 1995. **11**(1): p. 40-4.
232. Kaufman, C.K., et al., *GATA-3: an unexpected regulator of cell lineage determination in skin.* Genes & development, 2003. **17**(17): p. 2108-22.
233. Kouros-Mehr, H., et al., *GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland.* Cell, 2006. **127**(5): p. 1041-55.

234. Nichols, K.E., et al., *Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1.* Nature genetics, 2000. **24**(3): p. 266-70.
235. Freson, K., et al., *Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation.* Blood, 2001. **98**(1): p. 85-92.
236. Freson, K., et al., *Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation.* Hum Mol Genet, 2002. **11**(2): p. 147-52.
237. Mehaffey, M.G., et al., *X-linked thrombocytopenia caused by a novel mutation of GATA-1.* Blood, 2001. **98**(9): p. 2681-8.
238. Wechsler, J., et al., *Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome.* Nature genetics, 2002. **32**(1): p. 148-52.
239. Gamis, A.S. and J.M. Hilden, *Transient myeloproliferative disorder, a disorder with too few data and many unanswered questions: does it contain an important piece of the puzzle to understanding hematopoiesis and acute myelogenous leukemia?* J Pediatr Hematol Oncol, 2002. **24**(1): p. 2-5.
240. Lu, G., A.J. Altman, and P.A. Benn, *Review of the cytogenetic changes in acute megakaryoblastic leukemia: one disease or several?* Cancer Genet Cytogenet, 1993. **67**(2): p. 81-9.
241. Zipursky, A., A. Poon, and J. Doyle, *Leukemia in Down syndrome: a review.* Pediatr Hematol Oncol, 1992. **9**(2): p. 139-49.
242. Greene, M.E., et al., *Mutations in GATA1 in both transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down syndrome.* Blood Cells Mol Dis, 2003. **31**(3): p. 351-6.
243. Wieser, R., et al., *Transcription factor GATA-2 gene is located near 3q21 breakpoints in myeloid leukemia.* Biochem Biophys Res Commun, 2000. **273**(1): p. 239-45.
244. Fadilah, S.A., et al., *GATA-1 and GATA-2 gene expression is related to the severity of dysplasia in myelodysplastic syndrome.* Leukemia, 2002. **16**(8): p. 1563-5.
245. Ohyashiki, J.H., et al., *Ecotropic virus integration site-1 gene preferentially expressed in post-myelodysplasia acute myeloid leukemia: possible association with GATA-1, GATA-2, and stem cell leukemia gene expression.* Blood, 1995. **85**(12): p. 3713-8.
246. Vicente, C., et al., *Overexpression of GATA2 predicts an adverse prognosis for patients with acute myeloid leukemia and it is associated with distinct molecular abnormalities.* Leukemia, 2012. **26**(3): p. 550-4.
247. Yan, X.J., et al., *Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia.* Nature genetics, 2011. **43**(4): p. 309-15.
248. Seo, D., et al., *Gene expression phenotypes of atherosclerosis.* Arterioscler Thromb Vasc Biol, 2004. **24**(10): p. 1922-7.
249. Acosta, D., et al., *DNA methylation changes in murine breast adenocarcinomas allow the identification of candidate genes for human breast carcinogenesis.* Mamm Genome, 2011. **22**(3-4): p. 249-59.
250. Bohm, M., et al., *A role for GATA-2 in transition to an aggressive phenotype in prostate cancer through modulation of key androgen-regulated genes.* Oncogene, 2009. **28**(43): p. 3847-56.

251. Connelly, J.J., et al., *GATA2 is associated with familial early-onset coronary artery disease.* PLoS Genet, 2006. **2**(8): p. e139.
252. Scherzer, C.R., et al., *GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10907-12.

Chapter 2: Materials and Methods

2.1 Introduction

Materials and methods present in **Chapter 2** are used for experiments. Unless otherwise stated, materials and methods in section **2.1** and **2.2** apply to all chapters. Manufacturers, distributors and suppliers of the chemicals, kits and reagents used throughout the studies are quoted below.

1. Abcam - Sapphire Bioscience Pty Ltd, Waterloo, NSW, Australia.
2. Agilent - Integrated Sciences, Sydney, NSW, Australia.
3. Ambion® - Applied Biosystems, Scoresby, VIC, Australia.
4. Applied Biosystems - Life Technologies Australia Pty Ltd, Mulgrave, VIC, Australia.
5. BD Biosciences - BD Australia head office, North Ryde, NSW, Australia.
6. Bio-Rad - Bio-Rad Laboratories (Pacific) Pty Ltd, Gladesville, NSW, Australia.
7. Bio-Tek Instruments, Inc - Millennium Science, Mulgrave, VIC, Australia.
8. Cell Signaling Technology - Genesearch Pty Ltd, Arundel, QLD, Australia.
9. Clonetech - Scientifix Pty Ltd, Clayton, VIC, Australia.
10. GE Healthcare - GE Healthcare Bio-Sciences Pty Ltd, Rydalmere, NSW, Australia.
11. GeneWorks - GeneWorks, Hindmarsh, SA, Australia.
12. Invitrogen - Invitrogen Australia Pty Ltd, Mulgrave, VIC, Australia.
13. Kodak - Kodak, Sydney, NSW, Australia.
14. Millipore - Merck Pty Ltd, Kilsyth, VIC, Australia.
15. New England Biolabs (NEB) - Genesearch Pty Ltd, QLD, Australia.
16. Olympus - Olympus Australia Pty Ltd, Richmond, SA, Australia.
17. Peprotech - Abacus ALS Australia, Brisbane, QLD, Australia.
18. Pierce - Thermo Fisher Scientific, Scoresby, VIC, Australia.
19. Promega - Promega Corporation, Alexandria, NSW, Australia.
20. QIAGEN - QAGEN Pty Ltd, Doncaster, VIC, Australia.

21. Roche - Roche Diagnostics Australia Pty Ltd, Castle Hill, NSW, Australia.
22. Santa Cruz - Quantum Scientific, Murarrie, QLD, Australia.
23. Sigma - Sigma-Aldrich Pty Ltd, Castle Hill, NSW, Australia.
24. STEMCELL™ Technology - STEMCELL Technologies Australia Pty Ltd, Tullamarine, VIC, Australia.
25. Stratagene - Integrated Sciences Pty Ltd, Chatswood, NSW, Australia.
26. TAKARA - Scientifix Pty Ltd, Clayton, VIC, Australia.
27. Thermo Fisher Scientific - Thermo Fisher Scientific, Scoresby, VIC, Australia.

2.2 General methods

2.2.1 DNA Isolation

All plasmid DNA extractions were performed according to the manual of QIAGEN Plasmid DNA Mini or Midi Kits (QIAGEN).

2.2.2 Total RNA Isolation

Total RNA from cell lines were extracted using the RNeasy® Micro kit (QIAGEN) as described by the manufacturer's protocol. The quality and quantity of RNA were evaluated using the Nanodrop Spectrophotometer (Thermo Fisher Scientific) or the Agilent 2100 Bioanalyzer (Agilent).

2.2.3 DNA or RNA Purification

PCR products and restriction enzyme digested DNA fragments were purified using either QIAquick PCR Purification or Gel Extraction Kits (QIAGEN) as per manufacturer's protocol. In order to reduce genomic DNA contaminations, RNA samples were purged with DNA-free™ Kit (Ambion) as suggested by manufacturer's manual.

2.2.4 Restriction Enzyme Digestion

A list of restriction enzymes (REs) used in DNA digestion is stated in **Table 2.1**. All the reactions are performed in a PCR machine under the optimum conditions as suggested by their

respective manufacturers' protocols. The reaction mix was incubated at 37°C for 3h, then heat inactivated at 65° for 20min. Following restriction digestion, the DNA fragments were purified using QIAquick PCR Purification or Gel Extraction Kits (QIAGEN) as described in section **2.2.3**.

Table 2.1: List of Restriction Enzymes.

Restriction Enzyme	Company
*Bam*HI	NEB
*Bgl*II	NEB
*Eco*RI	Promega
*Nde*I	NEB
*Sal*I	NEB
*Xba*I	Promega
*Xho*I	Promega

2.2.5 DNA Dephosphorylation and Cloning

Regardless of the vectors, all 5' overhang ends generated by REs were dephosphorylated with antarctic phosphatase (NEB) as per manufaturer's manual. Next, the dephosphorylated vectors were purified using QIAquick PCR Purification kit (QIAGEN) as described in section **2.2.3**. Purified RE digested PCR products were then ligated to vector using T4 DNA ligase (NEB) according to the recommendation of the manufacturer.

After ligation, the ligated product was transformed into XL10-Gold Ultracompetent Cells (Strategene) using the heat-shock approach. Briefly, ligated product was gently mixed with XL10-Gold cells and incubated for 30 minutes on ice, heat-shock for 30s exactly at 42°C in a water bath and back on ice for additional 2 minutes. The transformed cells were then pre-cultured in LB (Luria Bertani) broth at 42°C for 2h before plating on LB agar containing 100μg/ml ampicillin (Sigma). The LB agar plates were incubated at 37°C for 16-18 hours. Recombinant clones were screened for correct insert size and orientation using the colony PCR approach as described in section **2.2.7**.

2.2.6 Routine PCR

Oligonucleotides used in PCR amplification and sequencing were designed using a web-based tool, Primer3 (http://frodo.wi.mit.edu/). Details of the primer sequence and their annealing

temperature were outlined in **Appendix A** to **C**. PCR amplifications were performed as per protocol of *Pfu*Ultra™ II Fusion HS DNA Polymerase (Agilent). Briefly, PCR reaction was performed in a 50µl volume containing 1µl of *Pfu*Ultra™ II fusion HS DNA polymerase, 1X concentration *Pfu*Ultra™ II reaction buffer, 10-30ng of dsDNA template, 200µM of dNTP mix, 0.2µM of each forward and reverse primers.

Unless specified, the PCR thermal cycling conditions are as below:

1.	Initial denaturation	95°C	30s	
2.	Denaturation	95°C	10 s	- repeat 2-4 for 25 cycles
3.	Annealing	60°C	30 s	- varying annealing temperature (Ta)
4.	Extension	72°C	1 - 3 min	- depending on the size of amplicon
5.	Extension	72°C	10 min	
6.	Cooling down	10°C	Forever	

2.2.7 Colony PCR

To rapidly screen ligation reactions for positive clones, a colony PCR approach was employed. PCR amplifications were performed as per protocol of AmpliTaq ® DNA polymerase (Applied Biosystems). Briefly, colonies are picked with sterile pipette tips and transferred into a 50µl volume PCR mix containing 0.03U AmpliTaq DNA Polymerase, 1X concentration PCR Buffer II, 200µM of dNTP mix, 2mM of $MgCl_2$, 0.2µM of each vector primer and gene specific primers. PCR thermal cycling conditions for colony PCR are similar to routine PCR details in section **2.2.6**.

2.2.8 Site-Directed Mutagenesis

A web-based tool, PrimerX (http://www.bioinformatics.org/primerx/) was used to design mutant primers (**Appendix B-2** and **C-2**). GATA2 allelic mutant series cDNAs (human and mouse) and mutated GATA binding sites in *Prox1* promoter/enhancer element (*PEE*) luciferase construct were generated by site directed mutagenesis as per protocol of Stratagene's QuikChange™ Site-Directed Mutagenesis kit (Clonetech). Briefly, each reaction was performed in a 50µl volume containing 0.05U of *Pfu*Ultra HF DNA polymerase, 1X concentration reaction buffer, 10ng of dsDNA template, 3µl of QuikSolution, 200µM of dNTP

mix, 120ng of each forward and reverse primer. The cycling parameters for the QuikChange Site-Directed Mutagenesis method are as below:

Step I PCR Amplification

1.	Initial denaturation	95°C	30 s	
2.	Denaturation	95°C	30 s	
3.	Annealing	55°C	1 min	
4.	Extension	68°C	5-8 min	repeat 2-4 for 16 cycles
5.	Cooling down	10°C	Forever	

Step II *Dpn*I Digestion

1.	Digestion	37°C	1 hour
2.	Cooling down	10°C	Forever

2.2.9 Tissue Culture

The human embryonic kidney cells, HEK293T & HEK293, and the monkey kidney fibroblast cells, Cos-7 were grown in DMEM media (Sigma) supplemented with 10% FBS (Sigma) and 2mM of L-glutamine in the presence of 10µg/ml gentamicin (Sigma). HEK293T cell lines were used for the production of lentiviral or retroviral stocks; whereas, HEK293 and Cos-7 cells were used for transient GATA2 protein expression studies. The promyelocytic leukaemia cell line, HL-60 and GATA2 WT and mutant transduced clones were maintained in RPMI media 1640 (Invitrogen) with 10% FBS and 10µg/ml gentamicin.

2.2.10 Lipofectamine-Mediated Transfection

All transfections of plasmid DNA vectors were carried out using the Lipofectamine™ 2000 Transfection Reagent (Invitrogen) as per manufacturer's protocols.

2.2.11 Whole Cell Lysate Preparation

Adherent cells and non-adherent cells were harvested using standard procedure. Cells were collected at 500×g for 5 min, washed 2X with chilled PBS prior to lysis in 200µl RIPA buffer (Sigma) containing cOmplete Protease Inhibitors (Roche) and PhosStop phosphatase inhibitors (Roche). The resultants were incubated on ice for 10 min, followed by centrifugation at 13,000rpm, for 10min at 4°C. The supernatants were transferred into fresh

microcentrifuges. Total protein concentrations were measured using Bradford Reagent (Bio-Rad) according to the manufacturer's protocol followed by colorimetric measurement using EL808 Ultra Microplate reader (Bio-Tek Instruments, Inc).

2.2.12 Nuclear Lysate Preparation

Nuclear lysates were prepared as described by Andrews and Faller [1]. Briefly, washed pelleted cells were re-suspened in 400µl cold Buffer A with protease and phosphotase inhibitors. The cells were incubated on ice for 10 min, following by vortexing for 10 seconds. Samples were centrifuged for 10 seconds, and the supernatant fraction was discarded. The pellets were lyses in 100-200µl Buffer C supplemented with cOmplete Protease Inhibitors (Roche) and PhosStop phosphatase inhibitors (Roche), and incubated on ice for 20 min. The samples were then spun for 2 min at 4°C and the supernatant fractions were transferred to new microcentrifuge tubes.

2.3 Chapter 3 Methods

Details methods of **Chapter 3** were discussed in a published paper and **Appendix A-2.**

2.4 Chapter 4 Methods

2.4.1 Generation of *GATA2* Mutant Constructs

Novel *GATA2* mutants: R361L, R362Q, C373R and R398W were generated by using site directed mutagenesis (see **Appendix B-2**, **Supplementary Table 1** for primers) as described in section **2.2.8**. The entire coding region of all mutants was Sanger sequenced to contain only appropriate changes in GATA2 sequence.

2.4.2 Luciferase Reporter Assay

Luciferase reporter assays were performed using Dual-Luciferase Reporter Assay System (Promega) using a GloMax®-Multi Detection System (Promega). The methodology is detailed in **Appendix A-2, Supplementary methods.**

2.4.3 Homology Modelling of GATA2 WT and Mutants

WT GATA2 was aligned with the template structure of murine GATA3 whose structure was obtained by X-ray crystallography [2]. Template identification alignment indicates GATA2 ZF2 is 96% (56/58) identical to the homologius murine GATA3 ZF3. Subsequently, the zinc finger 2 motif was subjected for homology modelling. The alignment mode of Swiss model workspace (http://swissmodel.expasy.org/) was used to predict the theoretical 3 dimensional structure of WT GATA2 [3]. For modelling of the mutants, WT residues were replaced with disease associated residues *in silico* by modifying the input primary sequences. The models were analyzed, finetuned and opportunely minimized using the Swiss PDB viewer.

2.4.4 Verification of GATA2 Theoretical Model

The final results were further verified using the online tools PROCHECK, WHAT_CHECK, ERRAT VERIFY 3D and PROVE on the Structure Analysis and verification Server at the WWW address http://nihserver.mbi.ucla.edu/SAVES/. The final structures were presented with Pymol 0.99rcs6 (Innocentive Product, Delano Scientific LLC.).

2.4.5 Generation of FLAG-*GATA2* Expression Constructs

FLAG-*GATA2* cDNAs (WT and mutants-coding region only) were PCR amplified from the pCMV6-XL-*GATA2* (WT or mutants) plasmids (as described in **2.4.1**) using the primers (Flag-GATA2-F and GATA2-R, see **Appendix B-2**, **Supplementary Table 2**) and *Pfu*Ultra™ II Fusion HS DNA Polymerase (Agilent), cloned into the unique *Xba*I site of pcDNA3.1. Colony PCR were performed as described in **2.2.7** using the pf5xUAS-F and Gata2 Ex3b r V2 primer pairs (**Appendix B-2**, **Supplementary Table 3**).

2.4.6 Generation of Inducible FLAG-*GATA2* Dual Lentiviral Expression Constructs

For generation of lentivirus constructs, *Xba*I digested FLAG-*GATA2* cDNAs derived from section **2.4.5** were cloned into pf5xUAS as described in section **2.2.5**. The vectors enable 4-hydroxytamoxifen (4-HT) inducible FLAG-tagged GATA2 expression and constitute puromycin expression. Colony PCR were performed as described in **2.2.7** using the pf5xUAS-F and Gata2 Ex3b r V2 primer pairs (**Appendix B-2**, **Supplementary Table 3**).

2.4.7 Generation of Regulatable FLAG-*GATA2* Expressing HL-60 Cell Lines

FLAG-*GATA2* expressing HL-60 cell lines were generated as described in **Appendix A-2**. Briefly, infectious third generation lentivirus was made by cotransfecting 293T cells with pf5xUAS-FLAG-*GATA2* (WT or mutants) and the three packaging plasmids pHCMVwhvgagpolml, pHCMV-G and pHCMVwhvrevml (kindly provided by Don Ansor). HL-60 containing Gal4-ERV (derived from section **2.3**) was then subsequently transduced with the FLAG-*GATA2* (WT and mutants) or EV lentiviral supernatant and selected in 3 μg/ml of Puromycin (Sigma).

2.4.8 Co-Immunoprecipitation

Nuclear lysates or whole cell lysates from HEK293, co-transfected with pcDNA3.1-FLAG-*GATA2* (WT or mutants) and pcDNA3-*PU.1* (isoform 2, NP_003111.2, kindly provided by Dan Tenen) were prepared as described in section **2.2.11** and **2.2.12,** and used for co-immunoprecipitation. The lysates were incubated in 500μl of lysis buffer with 2μg of anti-FLAG mouse monoclonal antibodies (M2, Sigma) at 4°C on a rotating wheel for an hour. Subsequently, 20μl Protein G-agarose beads (Roche) were added to each reaction and were further incubated for another hour. The beads were then washed extensively (3 times) in lysis buffer before resuspended in 2x sample loading buffer, boiled and resolved in SDS-PAGE. The pulled down proteins were eletrotransfered to PVDF membrane and analysed with polyclonal GATA2 (Santa Cruz) and monoclonal rabbit PU.1 (Cell Signaling) antibody. To pull down endogenous binding partner, nuclear lysates or whole cell lysates from induced HL-60 (with 100nM 4-HT) were prepared as mentioned above. Subsequently, the protein blot was analysed with monoclonal mouse PLZF antibody (Millipore).

2.4.9 Generation of *Gata2* Retrovirus Constructs

The cDNA of *Gata2* WT or mutants was PCR amplified (see **Appendix B-2**, **Supplementary Table 5** for primers) and cloned into pre-linearized (*Xho*I and *Eco*RI, NEB) pMSCV-IRES-GFP retroviral plasmids.

2.4.10 Generation of Retrovirus

Infectious retrovirus (ecotropic) was made by cotransfecting HEK293T cells with pMSCV-*Gata2*-IRES-GFP (WT or mutants) and the packaging plasmid pEQ-Eco (mass ratio 1:1). Supernatants were harvested 24h later and filtered through a 45 μm syringe filter (Thermo Fisher Scientific).

2.4.11 Animal Handling and Bone Marrow Cells Extraction

All experiments were performed according to protocols approved by the University of Adelaide Animal Ethics Committee (M-2012-027) and SA Pathology Animal Ethics Commitee (109/11). All mice used in the study were C57BL/6. Mice were kept under a 12h light/12h dark cycle with unlimited access to food and water, and were humanely scarified by cervical dislocation. Femurs and tibias of 8- to 12-week-old female C57BL/6 were aseptically harvested and crushed using mortar and pestle in IMDM media (Invitrogen) supplemented with 15% FBS (Sigma). Red blood cells were depleted using RBC lysis buffer (QIAGEN). Total BM cells were then centrifuged (400xg, 10min), washed and counted in a haemocytometer. Subsequently, total BM cells were resuspended in IMDM supplemented with 15% FBS and 10μg/ml each of the TPO, SCF and G-CSF (Peprotech), plated in retronectin (30 μg/ml, TAKARA) coated plates , and cultured for 24 hours.

2.4.12 Retrovirus Transduction and Colony Forming Unit Assay

Retrovirus supernatant (WT or mutants) was added to 6-well plates containing total BM cells (volume ratio, 1:1). Subsequently, the plates were spun at 37°C, 2500 RPM for 90 minutes. The cells were then placed in incubator for 4h before a second similar transduction. After 24 h, the virus media was removed and replaced with fresh full IMDM media and continue cultured for another 24h. GFP positive BM cells were sorted using BD FACSAria cell sorter (BD Biosciences). For colony forming unit assay, GFP positive BM cells were plated in 5000 cells/ml of Methocult medium (STEMCELL™ technology) in the presence of TPO, SCF and G-CSF. All assays were performed in triplicate. After 7 days of culture, colonies were enumerated and typed under the light microscope (Olympus). Replating assays were performed by diluting the colonies and methylcellulose medium with fresh IMDM, then

centrifuged (400xg, 10min), and resuspended in IMDM media supplemented with 2% FBS before inoculated into fresh Methocult medium.

2.5 Chapter 5 Methods

2.5.1 Generation of *Gata2* Mutant Constructs

A mouse *Gata2* cDNA clone (pCMV6-Entry-Gata2) was purchased from OriGene, and T354M, T355del, L359V, R361L, R362Q, C373R and R398W mutants (see **Appendix C-2**, **Supplementary Table 1** for primers) were generated as in section **2.2.8**.

2.5.2 Immunofluorescence Staining

HEK293 cells transfected with WT or *Gata2* mutant plasmids were stained and prepared as described in **Appendix A-2**.

2.5.3 Generation of *PEE* GATA Mutated Site Luciferase Reporter Plasmid

pGL4.12-*PEE* with WT GATA binding sites were mutated by site directed mutagenesis (see **Appendix C-2**, **Supplementary Table 2** for primers) as described in section **2.2.8**.

2.6 Media and Solutions

Values in parentheses are final concentrations.

LB Agar (1000ml)

Agar	15g	(1.5% w/v)
LB broth	Top up to 999ml	
Boil to dissolve agar.		
Ampicillin (100mg/ml)	1ml	(100μg/ml)

0.5M EDTA pH8.0 (100ml)

Diaminoethane tetraacetic acid	18.6g	(0.5M)
DEPC-water	Top up to 100ml	
pH to 8.0 using NaOH.		
Sterilised by autoclaving.		

1.5M Tris.HCl pH8.8 (150ml)

Tris base	27.23g	(1.5M)
Filtered water	80ml	
Adjust to pH8.8 with 1M HCl.		
Filtered water	Top up to 150ml	

0.5M Tris.HCl pH6.8 (100ml)

Tris base	6g	(0.5M)
Filtered water 60ml		
Adjust to pH6.8 with 1M HCl.		
Filtered water	Top up to100ml	

10% SDS (400ml)

Sodium dodecyl sulphate (SDS)	40g	(10% w/v)
Filtered water	Top up to 400ml	
Filter-sterilised with 0.2μm cartridge		

10% APS (6ml)

Ammonium persulfate (APS)	0.6g	(10% w/v)
Filtered water	Top up to 6ml	
Aliquots of 30μl and 50μl. Store at -20°C.		

10x running buffer for SDS-PAGE (1000ml)

Tris base	30.3g	(3% w/v)
Glycine	144g	(14.4% w/v)
SDS	10g	(1% w/v)
Filtered water	800ml	
pH to 8.3 with 1M HCl/1M NaOH		
Filtered water	Top up to 1000ml	

1X transfer buffer (4000ml)

Tris base	12.12g	(0.3% w/v)
Glycine	57.6g	(1.44% w/v)
Filtered water	3000ml	
Stir until dissolve.		
Methanol	800ml	(20% v/v)
Filtered water	Top up to 4000 ml	
Store at 4°C.		

PBS-T (1000ml)

Phosphate buffered saline (PBS)	999ml	
Tween-20	1ml	(0.1% v/v)

Blocking buffer for western blotting (1000ml)

Non-fat dried milk powder	50g	(5% w/v)
PBS-T	Top up to 1000ml	

10X TGE (Tris-glycine-EDTA) buffer (1000ml)

Trizma base	30.28 g	(0.25 M)
Glycine	142.7 g	(1.9 M)
EDTA	3.92 g	(10 mM)
pH to 8.3 with 1M HCl/1M NaOH		
Filtered water	Top up to 1000ml	

Buffer A

HEPES-KOH pH 7.9 (200mM)	50µl	(10mM)
MgCl2 (100mM)	15µl	(15mM)
KCl (1M)	10µl	(10mM)
Dithiothreitol (1M)	0.5µl	(0.5mM)
Roche complete protease inhibitors (100X)	10µl	(1X)
H2O	914.5µl	

Buffer C

HEPES-KOH pH 7.9 (200mM)	100µl	(20mM)
Glycerol (50%)	500µl	(25%)
NaCl (5M)	84µl	(420mM)
MgCl2 (100mM)	15µl	(1.5mM)
EDTA (20mM)	10µl	(0.2mM)
Dithiothreitol (1M)	0.5µl	(0.5mM)
Roche complete protease inhibitors (100X)	10µl	(1X)
H2O	280.5µl	

2.7 References

1. Andrews, N.C. and D.V. Faller, *A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells.* Nucleic acids research, 1991. **19**(9): p. 2499.
2. Bates, D.L., et al., *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, 2008. **381**(5): p. 1292-306.
3. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

Chapter 3: Novel *GATA2* Mutations in Familial MDS/AML

3.1 Summary

Acute myeloid leukaemia is a heterogeneous malignancy caused by a plethora of mutations that lead to clonal proliferation of immature haematopoietic progenitor cells (HPCs) in the peripheral blood or bone marrow. *RUNX1* and *CEBPA* are 2 hitherto known mutations associated with familial AML. In this study, we report the discovery of 2 novel *GATA2* mutations, T355del (p.Thr355del) and T354M (p.Thr354Met) in 4 separate families affected with MDS/AML. Genetic analyses indicate that these mutations segregated with the disease, and are not present in 1,390 normal chromosomes or 268 sporadic AML samples. Taken together, we report that *GATA2* is a new predisposition gene for MDS/AML.

To gain further insight into the role of *GATA2* mutation in MDS/AML, we performed a series of functional characterizations on these mutants. GATA2 wild type (WT), T354M, T355del and L359V (a recurrent mutation reported in CML-blast crisis; p.Leu359Val) constructs were transiently expressed in HEK293, respectively. All GATA2 proteins were appropriately localized to the nucleus. However, T354M exhibited reduced DNA binding potential to the GATA protein consensus WGATAR DNA motif; whereas, T355del bound very weakly if at all. Consistent with this, T354M and T355del showed significantly reduced GATA2 transactivation of known *GATA2* responsive elements in luciferase reporter assays. Of note, co-transfection of T354M or T355del with WT significantly reduced WT transactivation activity, suggesting that these mutants act in a dominant negative fashion. Conversely, L359V behaved almost similar or with enhanced activity compared to WT in all assays.

Given the expression GATA2 is predominantly in HSC or myeloid cells in the haematopoietic system, we generated 4 hydroxy tamoxifen (4HT) regulatable stably transduced promyelocytic

HL60 cell lines expressing WT, T354M, T355del and L359V GATA2 protein. Under non-differentiating conditions, WT and L359V inhibited cell growth; while T354M allowed proliferation. When induced with the strong differentiation agent, all *trans* retinoic acid (ATRA), WT and T354M inhibited differentiation of HL60 cells. Together, T354M allowed cell proliferation/survival, while simultaneously reducing apoptosis. In contrast, T355del had no effects in all assays, suggesting that it is a complete LOF mutation. Microarray analysis of the *GATA2* mutants in HL60 cells indicated that T354M and T355del at the gene transactivation level are predominantly loss-of-function. Notably, L359V displayed both a gain and partial loss-of-function. The landmark finding of this study is the identification of a new MDS/AML predisposition gene. In addition, we successfully characterized the molecular landscape of *GATA2* mutations that perturb its normal functions. In **Chapter 4** and **5**, we further characterized novel *GATA2* mutations associated with other disorders.

3.2 Notes

The published article and supplementary information file have been included as Appendix A

1. Statement of authorship
2. Supplementary information

3.3 Permission to Reuse Published Materials

Ownership of copyright in the article remains with the Authors, and provided that, when reproducing the Contribution or extracts from it, the Authors acknowledge first and reference publication in the Journal, the Authors retain the following non-exclusive rights:

a) To reproduce the Contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

b) They and any academic institution where they work at the time may reproduce the Contribution for the purpose of course teaching.

c) To reuse figures or tables created by them and contained in the Contribution in other works created by them.

d) To post a copy of the Contribution as accepted for publication after peer review (in Word or Text format) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the Journal article on NPG's web site (eg through the DOI).

3.4 The Published Article

Please refer to the next page for the following published article:

Heritable *GATA2* mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia

Christopher N. Hahn, **Chan-Eng Chong**, Catherine L. Carmichael, Ella J. Wilkins, Peter J. Brautigan, Xiao-Chun Li, Milena Babic, Ming Lin, Amandine Carmagnac, Young K. Lee, Chung H. Kok, Lucia Gagliardi, Kathryn L. Friend, Paul G. Ekert, Carolyn M. Butcher, Anna L. Brown, Ian D. Lewis, L. Bik To, Andrew E. Timms, Jan Storek, Sarah Moore, Meryl Altree, Robert Escher, Peter G. Bardy, Graeme K. Suthers, Richard J. D'Andrea, Marshall S. Horwitz & Hamish S. Scott.

"This research was originally published in *Nat Genet.* 2011 Sep 4;43(10):1012-7. doi: 10.1038/ng.913.© Nature Publishing Group."

Hahn, C.N., Chong, C., Carmichael, C.L., Wilkins, E.J., Brautigan, P.J., Li, X., Babic, M., Lin, M., Carmagnac, A., Lee, Y.K., Kok, C.H., Gagliardi, L., Friend, K.L., Ekert, P.G., Butcher, C.M., Brown, A.L., Lewis, I.D., To, L.B., Timms, A.E., Storek, J., Moore, S., Altree, M., Escher, R., Bardy, P.G., Suthers, G.K., D'Andrea, R.J., Horwitz, M.S. & Scott, H.S. (2011) Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia.
Nature Genetics, v. 43(10), pp. 1012-1017

3.5 Additional Results

Western blotting-Electrophoretic Mobility Shift Assay (WEMSA) is a simple, robust and highly reproducible semi-quantatitve method for measuring DNA binding strength of a DNA binding protein such as a transcription factor. To accurately determine DNA binding strength of GATA2 WT and mutants, we employed Isothermal Titration Calorimetry (ITC) in the present study. To do this, the region of the human GATA2 cDNA encoding ZF2 domain only (*i.e.* residues 328-409) was cloned and overexpressed from a bacterial expression system. ZF2 protein was purified using cation-exchange chromatography followed by size exclusion chromatography. The resultant protein was checked for folding using circular dichrosim and 1D H-NMR. We then test the ZF2 domain binding potential using oligonucleotide probes derived from the human *GM-CSF* promoter and murine *Prox1* promoter using Electrophoretic Mobility Shift Assay (EMSA, also known as gel shift assay). Our assays showed that GATA2 WT bound equally as a single bound species to both probes (**Figure 3.1a**). EMSA also showed that WT and L359V proteins had indistinguishable binding affinity. However, T354M showed a subtle but reproducible weaker affinity, while T355del showed a more obvious weaker affinity as indicated by the higher protein concentrations required to observe the appearance of the complex (**Figure 3.1b**).

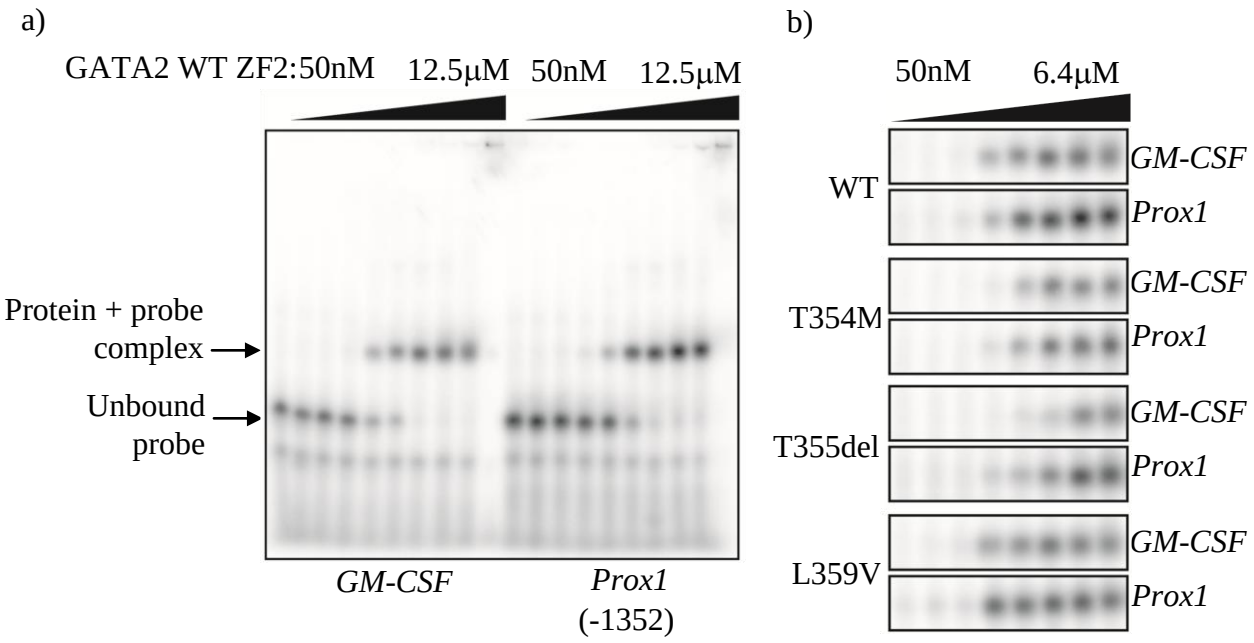


Figure 3.1: Binding of GATA2 ZF2 to the *GM-CSF* and *Prox1* Oligonucleotide Probes Analyzed by EMSA.

a) WT protein was titrated with human *GM-CSF* and murine *Prox1* -1352 (in 2 fold increasing concentration proteins) showing the disappearance of the unbound probe and the appearance of a single bound species (with protein/DNA complexes occurring at the higher protein concentration). b) EMSA study of GATA2 mutants proteins on *GM-CSF* and *Prox1* -1352 probe.

We then carried out the ITC experiments on the constructs, in which GATA2 ZF2 protein was titrated with *GM-CSF* probe. Our data showed that a high affinity binding event at a 1:1 (protein:oligonucleotides) stoichiometry, as well as a weaker event, which probably represents a non-specific binding of GATA2 (data not shown). Nevertheless, ITC results indicated that L359V does not have a significantly different affinity for *GM-CSF* compare to WT, while the T354M and T355del have a measurably weaker binding affinity (**Table 3.1**). Unfortunately, we were unable to determine K_a and K_d for the *Prox1* probe due to multiple (at least three) distinguishable binding events in our ITC experiment. This type of data cannot be fitted with the available binding models, but demonstrates that in contrast to the EMSA, under these

conditions there is a qualitative difference in the binding behaviour of the ZF2 domain to the two different oligo sequences.

ZF2 construct	K_a	K_d (average and range)
WT	6.0 E7 (±4.1 E7) 7.0 E7 (±2.5 E7) 1.1 E8 (±4.3 E7) 3.0 E7 (±3.1 E6)	15 nM (9-33)
L359V	6.4 E7 (±1.0 E7) 8.3 E7 (±3.8 E7)	14 nM (12-16)
T354M	3.7 E7 (±5.9 E6) 3.0 E7 (±7.1 E6) 4.7 E7 (±1.2 E7)	26 nM (21-33)
T355del	6.4 E6 (±9.6 E5)	156 nM

Table 3.1: Association Constants (K_a) and Dissociation Constants (K_d) Determined by Isothermal Titration Calorimetry (ITC) for Each GATA2 Constructs Using the *GM-CSF* Probe.

GATA2 ZF2 (200-250 μM) was titrated with the *GM-CSF* probe (20 μM) at 25 °C in 20 mM Tris (pH 7.9), 150 mM NaCl, 0.5 mM TCEP. The K_a values were obtained from individual experiments (with the error of each fit, data not shown), and the average and range (in brackets) expressed as a K_d. K_a and K_d are reciprocally related, in which K_a refers to the binding of a ligand to a protein, while K_d is defined as the dissociation or breakdown of a complex. The lower the value of K_d, the lower the concentration required for protein-ligand interaction. In other words, the protein (GATA2) has a higher affinity for the ligand (DNA) if the K_d value is low. Conversely, the higher the K_d value indicates the lower affinity of a protein to its ligand.

3.6 General Discussion

3.6.1 GATA2 is a New Predisposition Gene for MDS and AML

Considerable advances have been made in leukaemia research over the last few decades. Findings in leukaemia research reveal approximately two-thirds (65%) of acute leukaemia is characterized by non-random somatically acquired chromosomal abnormality [1]. Inheritable leukaemias and lymphomas are relatively rare compared to other cancers. Using a candidate gene targeting approach, we first describe the heterozygous GATA2 T354M mutation in 3 families affected with MDS/AML and a T355del lesion in 1 family [2]. Systematic analyses on multigeneration pedigrees revealed the predisposition of MDS and/or MDS/AML in each of the families. These GATA2 variants were not found in 695 normal individuals, nor were GATA2 coding mutations seen in 268 sporadic AML samples or germline samples from 35 other families affected with familial AML and lymphoid malignancies. Unfortunately, little or no good quality genetic material was available from most affected GATA2 predisposed family members as most of them died within two years of diagnosis. Haplotype analysis revealed that T354M arose at least twice in the three families, suggesting the mutation is not a single founder effect. Furthermore, a rare GTCT heterozygous deletion (4.2% in normal population) in the *GATA2* promoter (-460) was discovered in Pedigree 1 and 2, which also harboured an uncommon heterozygous "T" variant (rs45463801) in the *GATA2* 3'UTR. These data are consistent with Pedigree 1 and 2 deriving from a common founder but Pedigree 3 being unique.

Germline mutations of *RUNX1* and *CEBPA* are hitherto known predisposition TFs that lead to an autosomal dominant predisposition to familial MDS/AML [3, 4]. Mutations in these genes generally manifest an overt pre-leukaemic phenotype such as thrombocytopenia (*RUNX1*) [4] and eosinophilia (*CEBPA*) [5] in predisposed carriers before transforming to AML. On the other hand, *GATA2* mutated (heterozygous) families developed a rapid and high penetrance disease course that progressed relentlessly to death. The onset of disease occurred at a variety of ages, typically in young adulthood to early 40s. Astonishingly, there are still asymptomatic carriers in their late adulthood (>60 years old). Of these affected family members, they also

harboured an array of different somatically acquired chromosomal aberrations with monosomy 7, trisomy 8, and trisomy 21 being recurrent. *GATA2* mutations are likely to direct leukaemogenesis in the myeloid compartment where it is more highly expressed. Subsequently acquired somatic mutations cooperate toward MDS and/or AML and likely determine the particular subtype of the disease. However, it is unclear whether somatic mutations in these familial patients are driven by existing GATA2 mutation or whether randomly occurring mutations are selected for in a GATA2 mutant environment. We further demonstrated that while both mutants do not affect nuclear localization, they do alter DNA binding capacity to the WGATAR DNA motif in both WEMSA and ICT studies. We have also shown that the GATA2 mutations perturb transactivation of several target promoters. Of note, transient co-expression of mutant and WT protein (at a 1:1 ratio) displayed dominant negative activity, at least in some settings such as for the *CSF1R* promoter.

Assays in HL-60 cells looking at several "hallmarks of cancer" demonstrated clear differences between 3 GATA2 mutants tested. T354M was most remarkable enabling cell growth with reduced apoptosis even in the presence of the powerful differentiating agent, ATRA. T355del is a complete LOF mutant in these assays resembling the empty vector. The CML-BC mutant (L359V) mimicked WT protein in most assays. Collectively, our genetics data and *in vitro* functional assays indicate *GATA2* mutations are a conferral for cancerous properties.

3.6.2 Role of GATA2 in Haematological Disease

CML is a haematological malignancy with distinct biology and clinical features. Zhang *et al*. reported a gain of function mutation, L359V in the ZF2 domain of GATA2, which occurred in approximately 10% (8/85) of sporadic CML-BC cases, a disease often phenotypically akin to AML [6]. L359V is a specific somatic mutation that exclusively found in CML-BC, but not in other haematological malignancies (*i.e*. AML, ALL, MDS, CLL, BCR-ABL negative MPD or CML chronic phase) [7]. Both Thr354 and Thr355 residues do not contact DNA, but form polar contacts with each other via their side chains. Conversely, Leu359 is a hydrophobic residue that normally is buried in the DNA major groove and forms a close contact with DNA. L359V behaved similar to WT, except in DNA binding and, the differentiation and apoptosis

assays. We reason that pathogenic mutations leading to familial MDS and are likely to be caused by aberrant protein-protein interaction. GATA2 is known to form a transcriptional complex with one or more of its the interacting partners. Thereby, we cannot omit that the altered ZF2 domain caused by T354M and T355del may disrupt the formation of transcription complex. Regarding the dominant negative effect, we conjectured that these mutants may result in the aberrant protein-protein interaction either through the loss of an essential interaction or the formation of aberrant protein partnerships, eventually affecting transcription and translation of genes critical to myelopoiesis.

Consistent with other studies [6, 8], our data indicate *GATA2* mutations are not detected in a large heterogeneous cohort of sporadic AML patients. Yet, we cannot rule out that low percentage of blasts in AML samples may have masked the underlying mutation in these cells. The mechanism of GATA2 mutations in driving haematological malignancies is not clear. Nevertheless, the discovery of L359V may give us a lead in understanding the causative roles of mutant *GATA2* alleles in blood cancer. The pathogenic mechanism of Ph^+ CML is predominantly driven by BCR-ABL. As a consequence, *GATA2* is susceptible for further mutation caused by genomic instabilities, ultimately, leading to the blast crisis transformation [7]. In the hereditary MDS/AML, *GATA2* germline mutations are likely to initiate leukaemogenesis and later cooperate with somatic mutations for the disease development. We speculate that specific mutations in specific position in ZF2 of GATA2 together with other cooperative mutations ultimately give rise to the distinct phenotypic features of the disease. Consistent with this, somatic *GATA2* mutations in ZF1 and ZF2 domains (4/112) have been recently reported in AML-M5 [9], suggesting acquired *GATA2* mutations may correlate with or even drive specific subtypes of AML.

Familial MDS/AML with mutated *GATA2* is a rare form of haematological malignancies. To our best knowledge, only 4 affected families have been identified to date (2010). Nevertheless, overexpression of *GATA2* has been reported in many sporadic MDS[10], MDS/AML particularly in patients with 3q21q26 rearrangements [11, 12] and in *FLT3-ITD*$^+$ AML (GEO: GSE3505) [13], suggesting that the leukemogenic mechanism could be GATA2 deregulation

rather than direct mutation, such that upregulation of GATA2 might mirror the immature differentiation state of the AML type. In summary, the mechanism(s) by which GATA2 mutation or overexpression drive blood disorders such as MDS/AML are still unknown and remained to be solved.

3.7 Conclusion

Our study demonstrates that *GATA2* is a new predisposition gene for familial MDS/AML. Mutations within a highly conserved threonine string in ZF2 perturb GATA2 functions, highlighting the importance of GATA2 in normal haematopoiesis. The overall patient survival associated with MDS/AML harbouring these mutations is very poor, although there are heterozygous carriers over 60 years old. Nonetheless, our study implicates that *GATA2* has potential to be an important diagnostic parameter in genetic testing.

3.8 References

1. Look, A.T., *Oncogenic transcription factors in the human acute leukemias.* Science, 1997. **278**(5340): p. 1059-64.
2. Growney, J.D., et al., *Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype.* Blood, 2005. **106**(2): p. 494-504.
3. Smith, M.L., et al., *Mutation of CEBPA in familial acute myeloid leukemia.* The New England journal of medicine, 2004. **351**(23): p. 2403-7.
4. Song, W.J., et al., *Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia.* Nature genetics, 1999. **23**(2): p. 166-75.
5. Sellick, G.S., et al., *Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia.* Leukemia, 2005. **19**(7): p. 1276-8.
6. Zhang, S.J., et al., *Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia.* Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2076-81.
7. Wang, Q., et al., *Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis.* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3444-9.
8. Zhang, S.J., J.Y. Shi, and J.Y. Li, *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.* Leuk Res, 2009. **33**(8): p. 1141-3.

9. Yan, X.J., et al., *Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia.* Nature genetics, 2011. **43**(4): p. 309-15.
10. Fadilah, S.A., et al., *GATA-1 and GATA-2 gene expression is related to the severity of dysplasia in myelodysplastic syndrome.* Leukemia, 2002. **16**(8): p. 1563-5.
11. Lahortiga, I., et al., *Molecular heterogeneity in AML/MDS patients with 3q21q26 rearrangements.* Genes Chromosomes Cancer, 2004. **40**(3): p. 179-89.
12. Wieser, R., et al., *Transcription factor GATA-2 gene is located near 3q21 breakpoints in myeloid leukemia.* Biochem Biophys Res Commun, 2000. **273**(1): p. 239-45.
13. Bullinger, L., et al., *Identification of acquired copy number alterations and uniparental disomies in cytogenetically normal acute myeloid leukemia using high-resolution single-nucleotide polymorphism analysis.* Leukemia, 2010. **24**(2): p. 438-49.

Chapter 4: Characterization of Novel *GATA2* Mutations Associated with Haematological Malignancies, Immunodeficiency Syndrome and Lymphoedema

4.1 Introduction

Haematopoiesis represents a continuum of differentiation, which requires the expression of lineage-specific transcription factors (TFs). Many of these TFs participate actively in gene regulation in a combinatorial manner, either to activate or to repress the expression of lineage-specific downstream target genes. In the haematopoietic compartment, GATA2 is expressed in haematopoietic stem/progenitor cells (HSPCs) and playing critical roles in survival and proliferation of these cells. In chapter 3, we reported the discovery of 2 novel GATA2 germline mutations (T354M and T355del) in 4 separate families affected with familial MDS/AML. Our data suggest that both T354M and T355del are loss-of-function (LOF) mutations with dominant negative characteristics in some settings.

Recently, other genetic lesions in GATA2 have been reported in AML-M5 (somatic), MonoMac/DCML deficiency syndrome (germline) and Emberger syndrome (MDS/AML and lymphoedema; germline). These findings implicate that germline and somatic GATA2 mutations have quite different clinical presentations and represent 2 separate entities. We further investigated the functions of a selected GATA2 allelic mutant series representing of major disease phenotypes: T354M (MDS/AML), T355del (MDS) [1], L359V (CML-BC) [2], R361L and C373R (Emberger syndrome) [3], R362Q (AML-M5 and bi*CEBPA* AML) [4] and R398W (MonoMac/DCML deficiency syndrome) [5, 6] to determine whether these might be phenotypic differences that could explain any clinical differences.

We studied the functions of these GATA2 disease causing mutants using *in vitro* tissue culture assays, proteomic strategies and *ex vivo* clonogenic assays. We showed that these GATA2 mutants did not overtly affect nuclear localization. However, all mutants demonstrated lower

GATA consensus DNA binding which reflected their reduced transactivation ability *in vitro*, except for R398W which showed weaker DNA binding but greater transactivation, and L359V which was higher in both assays. Intriguingly, several of the GATA2 mutants bound with different affinities to GATA binding sites in different gene promoters and enhancers, providing evidence for mutation-specific conferral of phenotype. Furthermore, we also present the evidence that GATA2 mutant proteins have the capacity to physically alter protein-protein interaction; particularly T354M and C373R were shown to have a stronger affinity for PU.1 compared to WT. Our data demonstrated that, in addition to altering DNA binding, GATA2 mutants also influenced the interaction with key binding partners. Moreover, our pilot clonogenic experiment revealed that forced expression of GATA2 WT inhibited colony formation of HPCs. In contrast, GATA2 mutants appear to be LOF and do not confer self-renewal capacity to the HPCs, but allowed myeloid progenitor differentiation *ex vivo*. We conclude that each mutation confers specific but subtle changes to GATA2 function.

4.2 Results

4.2.1 GATA2 Mutants Appropriately Localize to the Nucleus

GATA2 is a transcription factor predominantly localized in the nucleus compartment. To ascertain whether GATA2 mutations affect nuclear localization, R361L, R362Q, C373R and R398W constructs were generated using site-directed mutagenesis. These constructs were individually transfected and expressed in HEK293 fibroblast cells. In line with the WT, all mutants appropriately target the nucleus, implicating that these mutants do not alter nuclear localization.

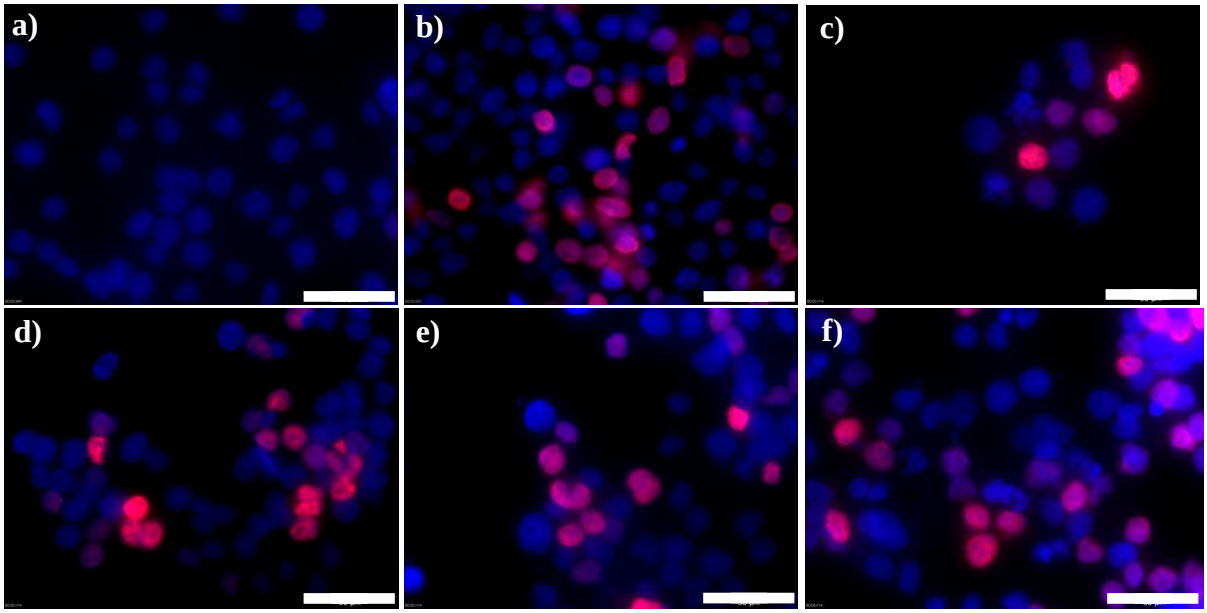


Figure 4.1: Expression of GATA2 Mutants in HEK293 Cells.

Immunofluorescence staining of GATA2 using anti-GATA2 antibodies and Alexa 555-labelled secondary antibodies. The cells were stained for GATA2 (pink) and DAPI (blue). Scale bars, 50 μm, 400X magnification. a) Empty vector (pCMV6-XL6), b) GATA2 WT, c) R361L, d) R362Q, e) C373R and f) R398W. Identical nuclear localization results were also obtained for the same mutant allelic series for murine GATA2 mutant proteins (see **Appendix C-3, Supplementary Figure 1).**

4.2.2 GATA2 Mutants Reduce Transactivation

In order to investigate whether or not GATA2 mutations have effects on its biological function, we performed luciferase reporter assay. This assay provides a quantitative measure for each of the mutants on their individual and/or combinatory contribution to transactivate known downstream target genes. [1]. The *CD34* enhancer and *LYL1* promoter contain GATA2 responsive elements and these regions were cloned into luciferase construct. **Figure 4.2** shows the luciferase activity (fold change relative to EV) resulting from GATA2 overexpression in the presence of (a) CD34 enhancer or (b) *LYL1* promoter reporter construct. In accordance with our previous work (see **Chapter 3**), T354M and T355del exhibited reduced luciferase activities, while L359V showed enhanced transactivation in *CD34* enhancer (**Figure 4.2a**). In contrast, both Emberger syndrome mutants (R361L and C373R) showed basal luciferase activity similar to EV, suggesting they are likely LOF mutants. R362Q demonstrated activity similar to T354M, implicating that R362Q is a partial LOF mutant. While the magnitude of transactivation was greater for *LYL1* promoter reporters construct, similar transactivation profiles were seen, except that R398W showed reduced activity on the *LYL1* promoter, but not for the *CD34* enhancer (**Figure 4.2b**). To mimic the heterozygous mutation, we cotransfected mutant constructs and WT or EV (1:1) in the presence of the *LYL1* promoter reporters construct in Cos-7 cells (**Figure 4.2c**). T354M, T355del and C373R reduced activity of the WT GATA2, suggesting either LOF of these mutants or some kind of dominant negative effects. In contrast, no reduced luciferase activity was observed in other mutants (R361L, R362Q and R398W) when cotransfected with WT. Luciferase activity for each of these combinations is almost similar to WT alone, implicating the activity seen was probably conferred by WT.

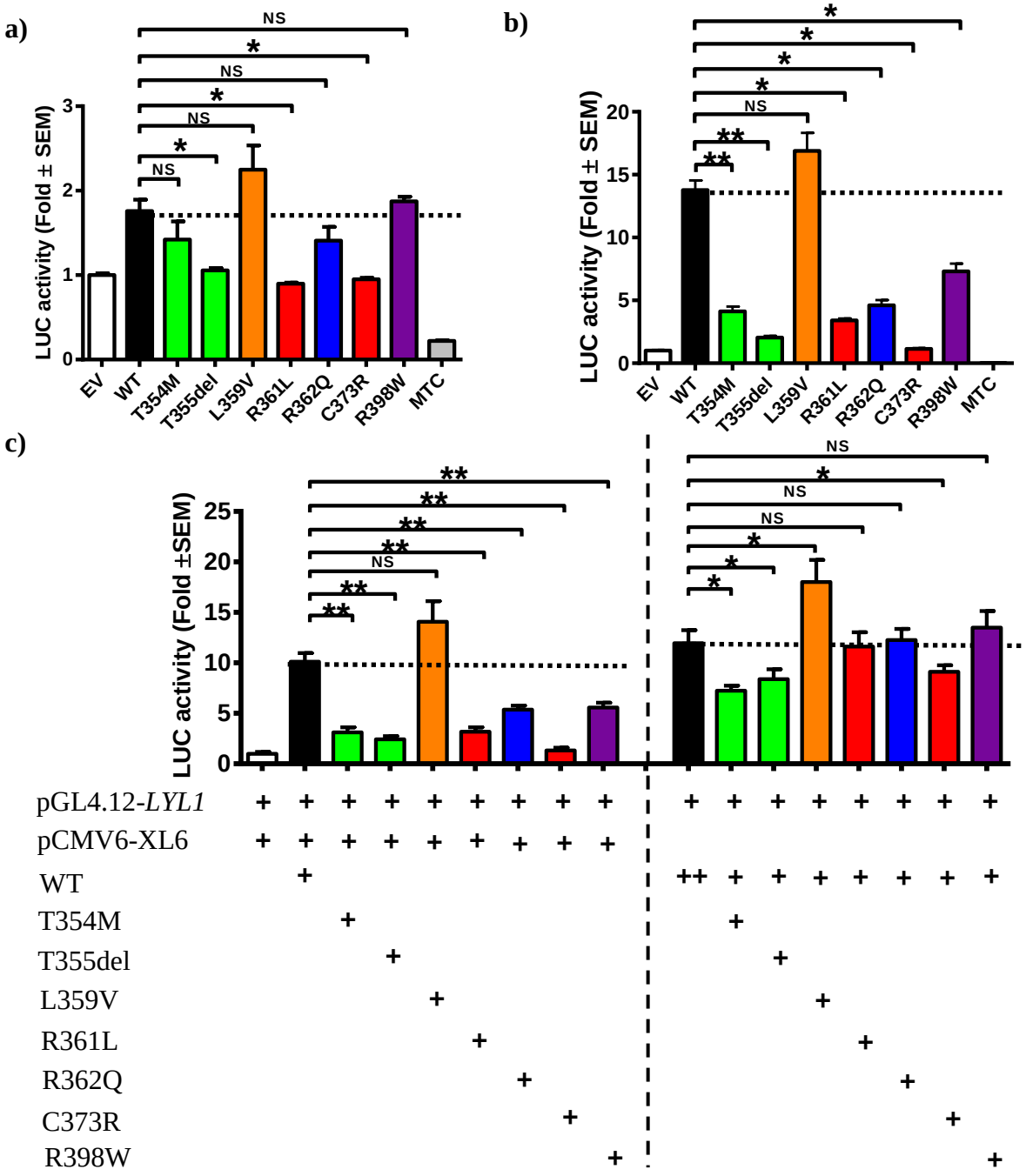


Figure 4.2: Transactivation of GATA2 on Known Responsive Elements.

a) **Transactivation of GATA2 on *CD34* enhancer.** HEK293 cells were cotransfected with 1) GATA2-responsive *CD34* x2 enhancer elements linked to a LUC reporter, and 2) GATA2 (WT or mutants) expression constructs or pCMV6-XL6 empty vector (EV). b) **Transactivation of GATA2 on *LYL1* promoter.** Cos-7 cells were co-transfected using *LYL1* promoter LUC as reporter and a GATA2 expression construct. c) **Dominant negative effect of GATA2 mutants on WT**. GATA2 WT or mutant construct was transfected into Cos-7 in the presence of the pGL4.12-*LYL1* (left). pGL4.12-*LYL1*, GATA2 WT and a mutant construct were cotransfected into Cos-7 (right). **P* < 0.05, **P* < 0.01, NS: not significant; MTC: mock transfection control. Green: MDS/AML; Orange: CML-BC; Red: Emberger syndrome; Blue: AML-M5 and Purple: MonoMac/DCML deficiency syndrome.

4.2.3 GATA2 Mutants Reduce DNA Binding Potential

To validate and to confirm the significance of conservative residues for GATA2 DNA binding, we performed WEMSAs on each of the mutant proteins. Nuclear lysates were harvested, quantified and subjected for DNA binding assays (**Figure 4.3a**). Human *TCRD* enhancer oligonucleotide [7] (contains a dual GATA recognition sites) was used as probe and was allowed to bind to our mutant allelic series of GATA2 nuclear lysates. WEMSA revealed a single GATA2-DNA complex with WT (**Figure 4.3b,** lane 3), but this was abolished using an oligonucleotide in which the 2 GATA sites were mutated from GATA to CTTA (lane 13). The specificity of the complex was confirmed by two positive controls (GATA consensus and granulocyte-macrophage colony stimulating factor (*GM-CSF* promoter), lane 11 and 12), which gave retarded bands of approximately the same size [8]. WEMSA of the mutant allelic series indicated that GATA2 mutant proteins exhibited differential DNA binding properties. As expected, the DNA binding for T354M, T355del and L359V was similar to our previous report [1], while DNA binding was abolished or greatly reduced in the Emberger syndrome mutants (R361L and C373R). In contrast, R362Q (sporadic AML-M5 and bi*CEBPA* AML) and R398W (germline and sporadic MonoMac/DCML syndrome) bound weakly compared to WT. The result indicates that these arginine residues are crucial to the ZF2 structure to enable DNA binding, whereas, C373R is a mutant with a severely disrupted ZF2 motif (Zn^{2+} coordinating 4 cysteines disrupted and one cysteine replaced by positively charge arginine to further repel zinc ion). Generally, the DNA binding affinity of GATA2 is summarized in the following order: L359V≥WT>R362Q/R398W>T354M>T355del/R361L/C373R, on the motifs tested, although this can vary substantially depending on the sequence of the WGATAR and the surrounding base pairs (see **Chapter 5**).

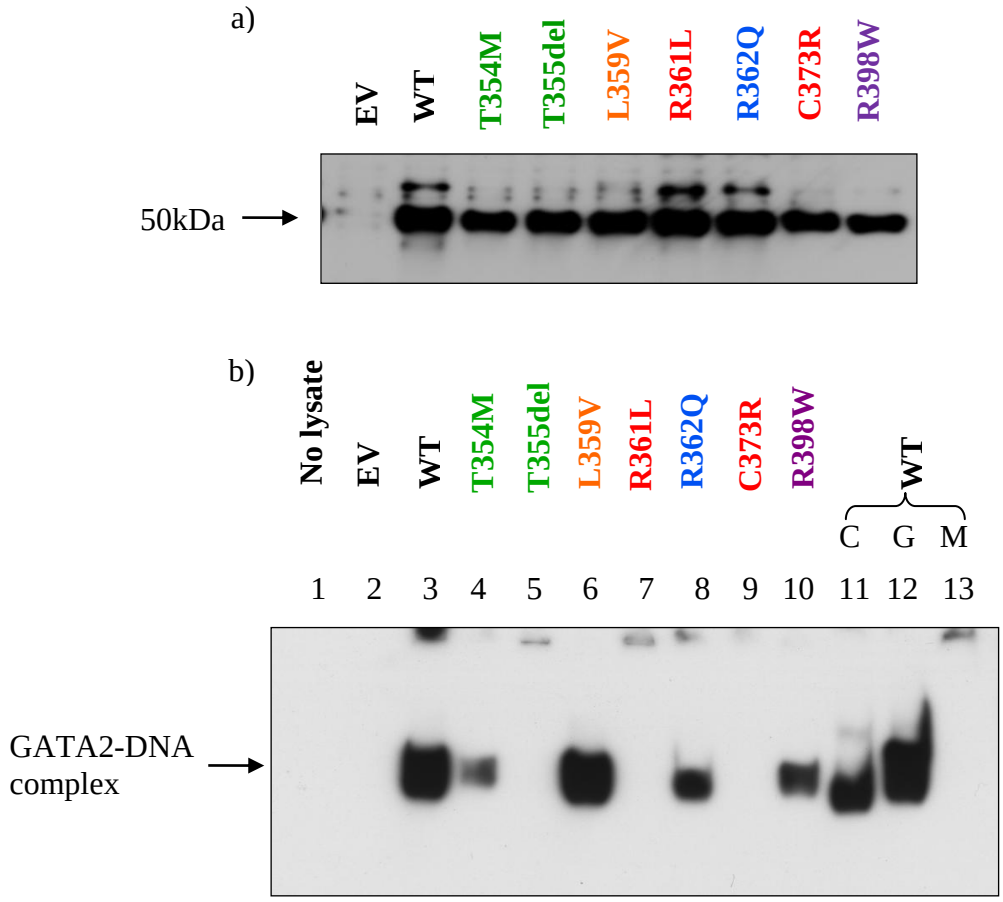


Figure 4.3: DNA Binding Assay on GATA2 WT and Mutant Proteins.

HEK293 cells were transfected individually with pCMV6-XL6 empty vector (EV), GATA2 WT or mutant expression construct. a) The resultant nuclear extracts (loading control) were used for Western analysis with anti-GATA2 antibody. b) 10µg each of the nuclear lysate was mixed with human *TCRD* enhancer oligonucleotide probe and then used for DNA binding assay. Oligonucleotide probes, C, G and M were used as control. C: GATA consensus; G: *GM-CSF* promoter and M: GATA binding site mutated *TCRD* enhancer. Note that a similar DNA binding pattern was found for murine GATA2 WT and mutant proteins (see **Chapter 5**).

4.2.4 Homology and Structural Modelling of GATA2 and Mutants

Arginine residues are highly conserved among vertebrate orthologs of GATA2 and are positioned along the ZF region. These residues have also been conserved throughout evolution within the GATA family members (**Figure 4.4**). All GATA family members share arginines at the positions 361, 362, 396 and 398 (relative to GATA2). To better understand the effect of amino acid substitutions in GATA2 protein on DNA binding, we carried out homology modelling on GATA2 ZF2 motif. Human GATA2 ZF2 motif (C-terminal) is 97% identical to murine GATA3 ZF2. Therefore, the GATA3-DNA crystal structure of the murine (PDB ID: 3DFX B chain) was used to model WT GATA2 and the mutant proteins [9]. **Figure 4.5** represents the theoretical structure of the WT GATA2 ZF2 and mutant proteins bound to DNA. Substitutions of residues at the mutation sites do not seem to change the overall structure of the GATA2 protein, except for C373R which is predicted to severely disrupt ZF2 domain. Our WT model demonstrates that the ZF2 motif forms a C-shaped clamp that interacts with DNA through the major groove of the core WGATAR consensus sequence. The replacement of the positively charged arginines (R361, R362 and R398 (red)) disrupts critical interactions (yellow dotted lines) not only to the adjacent residues, but also to the DNA. Perhaps, R361 is the most important residue in the ZF2 motif. The protonated side chain of R361 inserts deeply into the major groove of WGATAR sequence, interlocking a thymine residue (sense strand) and a cysteine residue (antisense strand) via hydrogen bonds. Substitution of arginine to a hydrophobic amino acid such as leucine significantly disrupts its binding to the WGATAR motif and intra molecular side chain-side chain interaction. On the other hand, R362 forms hydrogen bonds with the negative charged phosphate backbone near the DNA minor groove. Replacement with a polar uncharged glutamine is predicted to reduce DNA binding to GATA2. Similarly, an arginine to tryptophan mutation in residue 398 will likely to decrease DNA binding activity. In consistent with WEMSA and transactivation assays, our structural modelling showed that mutation of these key DNA binding residues is detriment GATA2 function, in particular, its affinity for DNA.

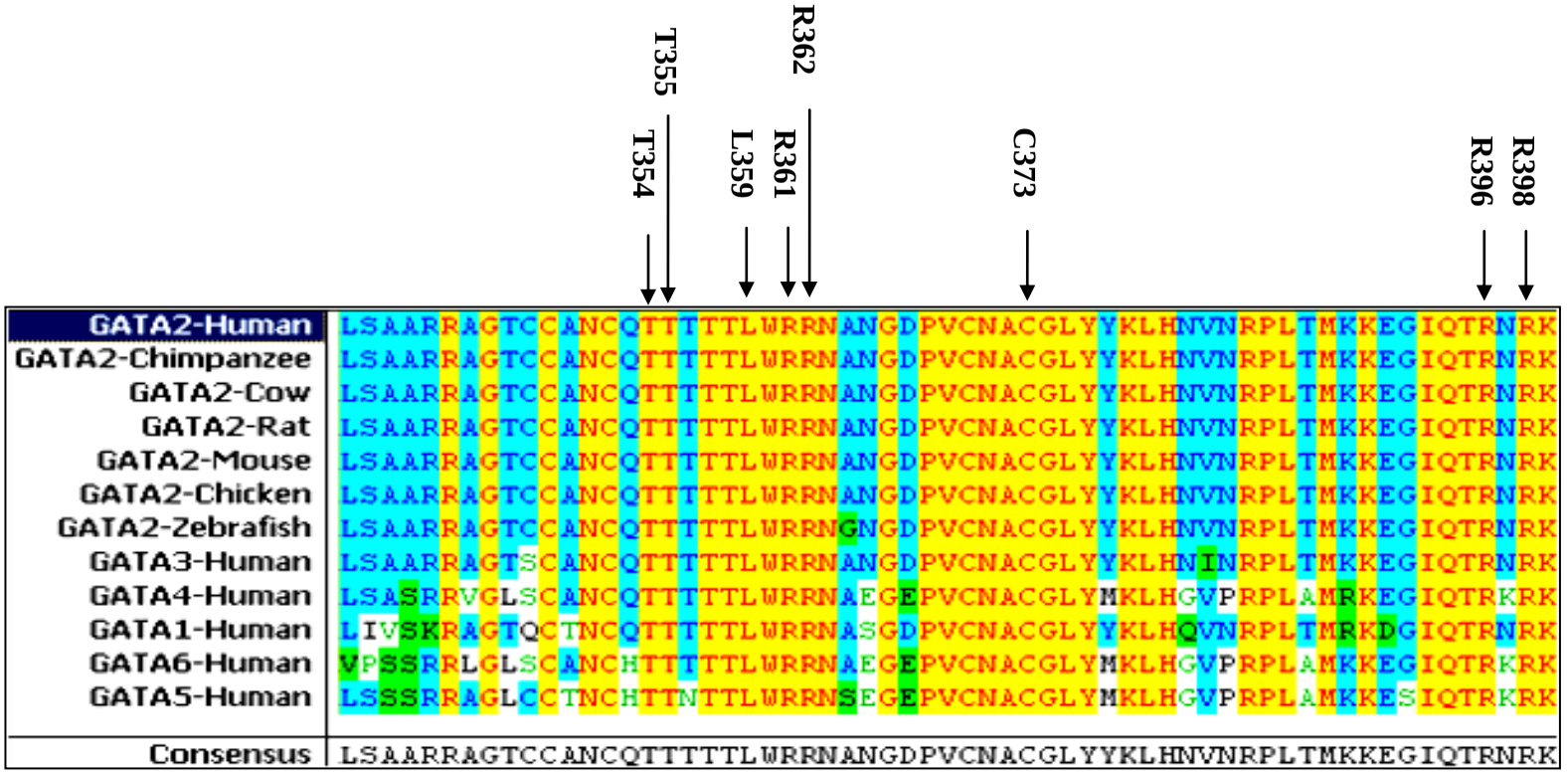


Figure 4.4: Multiple Sequence Alignment of GATA2 ZF2 Motif Across Different Vertebrates and Among GATA Family Members.

Residues such as arginines and cysteines (highlighted in yellow) are highly conserved in the GATA ZF2 motif region. Mutation of these highly conserved residues is predicted to be deleterious.

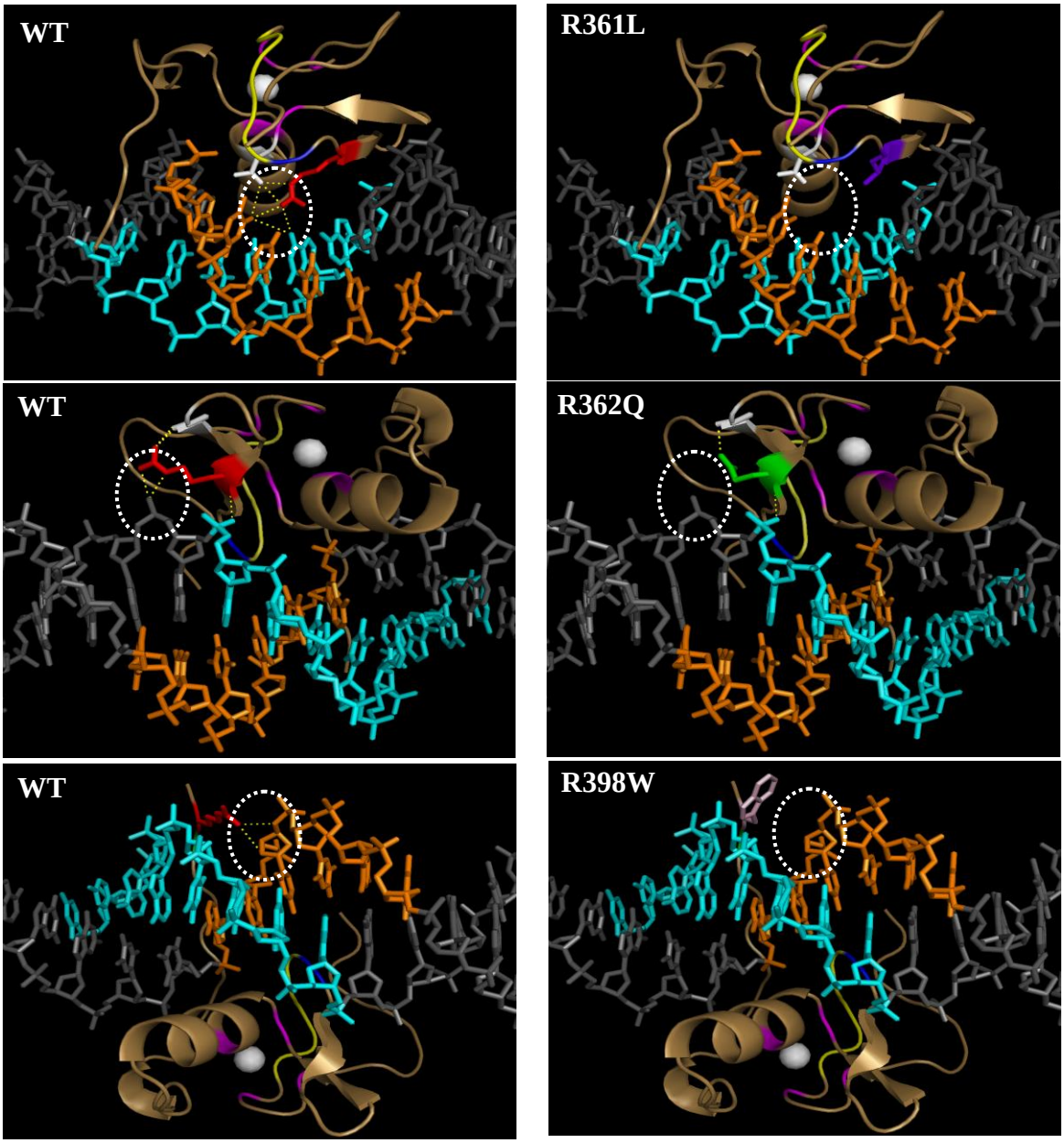


Figure 4.5: Structural Modelling of GATA2 WT and Mutants.

Homology modelling was performed to predict DNA binding potential of human GATA2 mutants by using mouse GATA3 (PDB ID: 3DFX B chain) as template. Abbreviation: The four zinc coordinating cysteines (magenta), zinc ion (white sphere), five consecutive threonines (T354-T358; yellow), L359 (blue), WT residues (red), adjacent interacting residues (white), R361L (purple blue), R362Q (green), R398W (light pink), major groove (AGATAA sense strand, cyan; TTATCT antisense strand, orange) and critical interactions (yellow dotted lines).

4.2.5 GATA2 Mutants Alter Protein-Protein Interaction

Besides acting as DNA binding domains, GATA2 ZF1 and ZF2 also function as active interacting sites for protein-protein interaction. TFs such as PU.1 interact with GATA2 through the ZF2 [10]. To determine whether GATA2 mutations alter protein-protein interaction, Co-IP experiments using the HEK293 transient transfection system were conducted. Transient transfections were performed by cotransfecting expression plasmids for PU.1 (isoform 2; NP_003111.2) and FLAG-tagged GATA2 (WT and mutants) or EV. After 24h, whole cell extracts (800μg each) were prepared and subjected to immunoprecipitation using anti-FLAG antibody. The immunoprecipitated proteins were then resolved in SDS-PAGE and electo-transferred to PVDF membrane before probing with anti-GATA2 antibody or anti-PU.1 antibody (**Figure 4.6**). Our results showed that PU.1 only co-precipitated with the presence of FLAG-GATA2 (WT or mutants), but not in the EV-transfected lysates, or in the no antibody control. Of particular note, T354M and C373R showed an enhanced affinity to PU.1 compared to WT. The enhanced interaction of GATA2 mutants with PU.1 is likely due to increased stability of the resulting GATA2-PU.1 complex. This experiment was repeated 3 times in total on independently prepared protein extracts and demonstrated stronger binding for T354M and C373R consistently.

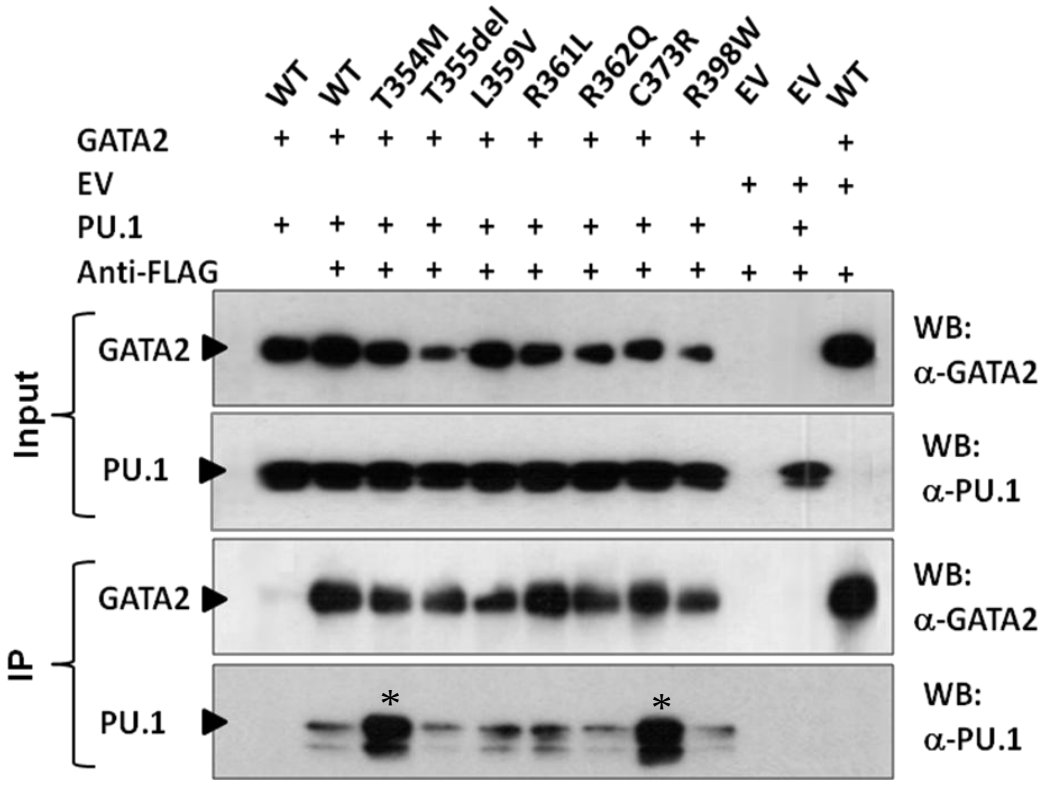


Figure 4.6: Co-Immunoprecipitation of GATA WT and Mutant Proteins and PU.1.
Whole cell lysates of HEK293 cells transfected with expression plasmids for FLAG-tagged GATA2 (WT and mutants) and PU.1 (isoform 2) were immunoprecipitated with anti-FLAG antibody. The specificity of Co-IP was confirmed by substitution of GATA2 WT and/or PU.1 expression plasmids with EV. 10μg of each lysate was used as an input control. Western blot (WB) analysis was performed with anti-GATA2 or anti-PU.1 antibody. T354M and C373R demonstrated altered binding partner activity (indicated by *).

4.2.6 GATA2 Mutants Allow Myeloid Progenitor Differentiation *Ex Vivo*

To gain further insight into the *in vivo* function of these human GATA2 mutants, we generated the same mutant allelic series for the mouse as was done for the human protein. We investigated the effect of *Gata2* WT and selected some representative mutants, *i.e.* T354M and L359V (partial LOF and some GOF), and Emberger syndrome mutants (complete LOF) for haematopoietic cell growth and progenitor differentiation assays. Total bone marrow (BM) cells were extracted from 8- to 12-week-old female C57BL/6 and transduced with retrovirus constitutively expressing *Gata2* (WT or mutants) and GFP (translated by an internal ribosomal entry site). $5x10^3$ GFP positive cells were mixed with MethoCult® GF M3434 medium and seeded in dishes for colony formation. This medium contains a cocktail of cytokines and growth factors that enables multilineage proliferation and colony formation particularly myeloid lineages. As shown in **Table 4.1**, colonies generated by GFP positive BM cells are predominantly CFU-G, followed by CFU-M. Relatively fewer bilineage CFU-GM, BFU-E and mixed-lineage colonies (CFU-GEMM) were obtained in individual experiments. This is consistent with other reports for whole BM studies [11].

The total colonies for WT (10.3 ± 1.9) were markedly less than EV control (54.7 ± 4.3; only 19% of EV control). However, the total colonies for all of the mutants (72.0 ± 2.9 to 108 ± 4.4 colonies) were higher than that from the EV control (54.7 ± 4.3) infected cells. This suggests that overexpression of *Gata2* WT inhibits proliferation of BM progenitor cells. Intriguingly, expression of GATA2 mutants significantly altered CFU-G and CFU-M colony formation with respect to WT. All mutants failed to suppress CFU-G as seen for WT GATA2 but rather increase colony formation (see **Figure 4.7a**). In contrast, T354M and L359V suppressed CFU-M at a level close to WT GATA2, while the two lymphoedema mutants (R361L and C373R) were similar to EV (see **Figure 4.7b**). This result is consistent with our previous functional studies, in which T354M and L359V are partial LOF (see **Chapter 3**, assays for HL-60 cells), while R361L and C373R are complete LOF (see **Chapter 5**). Together, our data suggest that each group of mutants confers different yet specific changes to HSPCs.

Table 4.1: Colony-Forming Activities of GFP-Sorted BM Cells.

GFP-positive BM cells expressing GATA2 (WT or mutants) or EV control were assayed for colony-forming activities (triplicate) in multiple myeloid conditions. Data are presented as mean with the SEM.

	Colony-forming units (CFU) per $5x10^3$ cells					
BM cells	CFU-G	CFU-M	CFU-GM	BFU-E	CFU-GEMM	Total
EV	33.3 ± 1.2	13.7 ± 3.4	5.3 ± 0.3	1.0 ± 0.6	1.3 ± 0.3	54.7 ± 4.3
WT	4.3 ± 2.0	5.3 ± 3.1	0.7 ± 0.7	0	0	10.3 ± 1.9
T354M	64.3 ± 3.4	6.7 ± 3.8	1.0 ± 0.6	0	0	72.0 ± 2.9
L359V	72.3 ± 5.5	3.3 ± 1.9	5.0 ± 0.6	0.7 ± 0.7	0.7 ± 0.3	82.0 ± 6.8
R361L	103.3 ± 6.7	18.0 ± 10.4	4.3 ± 0.7	0.7 ± 0.3	0.3 ± 0.3	126.7 ± 4.3
C373R	88.3 ± 4.4	16.3 ± 9.4	2.7 ± 0.7	0.3 ± 0.3	0.3 ± 0.3	108.0 ± 4.4

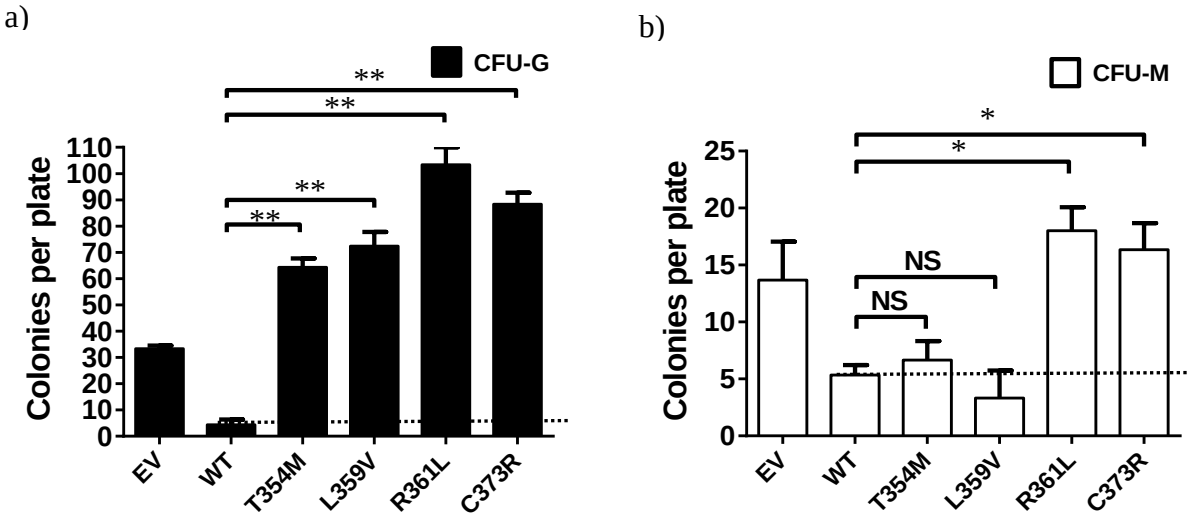


Figure 4.7 Whole BM Clonogenic Assays.

a) Whole BM cells transduced with mutant GATA2-expressing retrovirus do not suppress CFU-M as does WT GATA2. b) WT, T354M and L359V suppress CFU-M, while Emberger mutants (R361L and C373R) are complete LOF and similar to EV. **P<0.05*; ***P<0.01*; NS, not significant.

4.2.7 GATA2 Mutants Do Not Confer Self-Renewal Capacity *Ex Vivo*

To analyse the oncogenic effect of GATA2 mutations on BM cells, a serial replating assay was employed in our study, whereby the proliferative capacity and clonogenicity of progenitor cells were assessed in methylcellulose assays [12]. Normal HPCs can be replated at least 3 times before their proliferative potential becomes exhausted [12]. **Figure 4.8a** illustrates total CFU observed in serial replating assays, while **Figure 4.8b-f** indicate individual component for each type of CFUs seen on the methylcellulose medium. As mentioned previously, CFU generated by GFP positive BM cells are predominantly CFU-M and CFU-G (**Figure 4.8b-c**). In contrast, no or fewer GEMM, CFU-GM and BFU-E were seen in our serial replating assays (**Figure 4.8d-f**). As illustrated in **Figure 4.8a**, retrovirus mediated overexpression of GATA2 WT inhibited proliferation which is consistent with our previous studies on HL-60 cells [1]. In contrast, GATA2 mutants significantly increased proliferation and multilineage differentiation in primary assays (1°). These 1° cells were then harvested, pooled and replated in fresh methylcellulose medium. As expected in the secondary assays (2°), most of the lines exhibited a reduction in total number of colonies, particularly in T354M and C373R which demonstrated a sharp drop in total CFU. Of note, GATA2 WT demonstrated a 2-fold increased of total CFU compared to the 1° assays, suggesting that the inhibition of differentiation is not complete. In the tertiary assays (3°), while normal BM cells (*i.e.* EV) could still survive and form colonies, GATA2 WT and all mutants produced very few colonies, probably due to exhaustion during 2° replating (**Figure 4.8d-e**). No colonies were formed upon the fourth round of replating, yet diffused and scattered differentiating cells were seen (data not shown). Lastly, no BFU-E were seen in our subsequent replating assays (2°and 3°; **Figure 4.8f**), suggesting the culture conditions may not be optimal for the growth of BFU-E. These data, although preliminary, suggest that *Gata2* mutants (at least alone) do not confer self-renewal capacity on committed BM progenitors. Moreover, GATA2 mutants did not overtly enhance the serial replating capacity of HSPCs *in vitro*, possibly require other genetic events to cooperate to produce a leukaemic phenotype. All results in section **4.2** are summarized in **Table 4.2**.

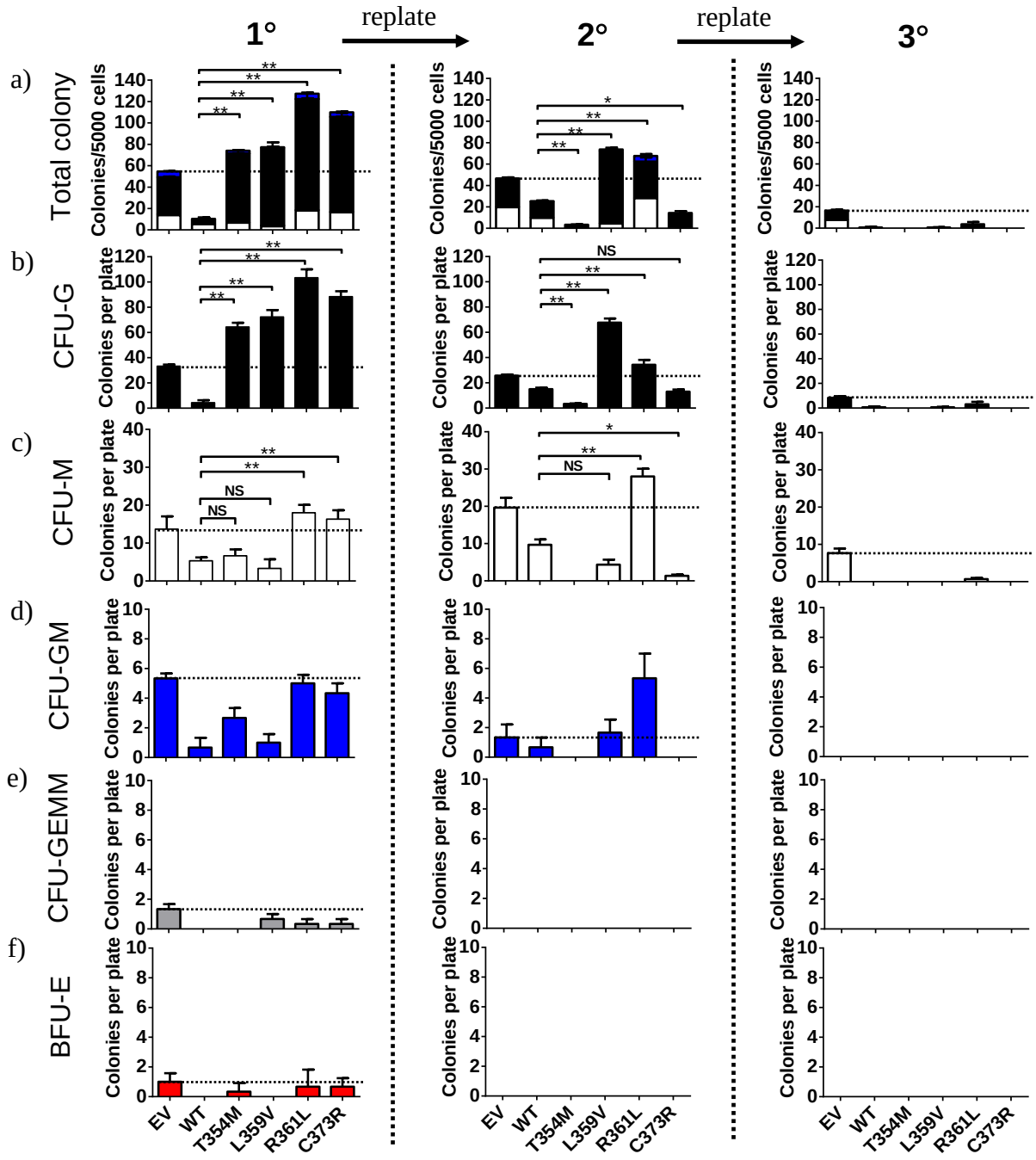


Figure 4.8: Clonogenic Assays on Total Bone Marrow Cells Transduced with *Gata2* WT or Mutants.

A serial replating was performed on retrovirally transduced mouse BM cells with either EV control or *Gata2* (WT or mutants). These cells were plated in methylcellulose medium (triplicate) supplemented with cytokines ($5x10^3$cells/dish) for 24h after the GFP selection (denoted as 1°). These cells were then harvested from 1° culture after 7days and were replated into new dishes at the same cell density (denoted as 2°). Tertiary colony assays (3°) were performed by replating cells derived from the 2° cultures. Colonies for each round of assays were enumerated and typed every 7 days. *$P<0.05$; **$P<0.01$; NS, not significant.

Table 4.2: Summary of the *In Vitro* Characterization of GATA2 Mutants.

Abbreviations: NL, nuclear localization; DB, DNA binding; TA, transactivation (promoter and cell type dependence*); PI, protein:protein interaction; Diff., differentiation; SR, self-renewal; N/A, not available. Empty vector (EV[#]) is used as vehicle control (basal activity) in all assays.

GATA2 Mutant	Assay					
	NL	DB	TA	PI	Diff.	SR
EV[#]	-	-	-	-	-	-
WT	✓	+++	+++	+	✘	✘
T354M	✓	+	++	++	✓	✘
T355del	✓	-	-/+*	+	N/A	N/A
L359V	✓	++++	++++	+	✓	✘
R361L	✓	-	-/+*	+	✓	✘
R362Q	✓	++	++	+	N/A	N/A
C373R	✓	-	-/+*	++	✓	✘
R398W	✓	++	+/+++*	+	N/A	N/A

4.3 Discussion

4.3.1 Substitution of Conserved Residues in GATA2 Disrupts Its Function

In the present study, we investigate the effect of clinically relevant *GATA2* mutations on its function. Our data suggest mutation on several critical residues severely damage GATA2 function. Subcellular localization studies demonstrated that missense mutations did not overtly alter GATA2 nuclear localization. However, GATA2 mutations perturbed transactivation of selected target promoters *in vitro*. Overall, most of the GATA2 mutations (except L359V in both assays and R398W on *CD34* enhancer) displayed predominantly LOF characteristics in luciferase reporter assays, which is consistent with DNA binding properties of the individual mutants. R398W is particularly interesting in which the mutant displays a context dependent transactivation activity. R398W was reported in MonoMac/DCML deficiency syndrome by 2 independent groups [5, 6]. Yet, no functional study was performed on them. The unique transactivation pattern of R398W may be reflected by its clinical phenotype as no haematological malignancies were reported in 6 patients harbouring R398W mutation [5, 6]. R398W may selectively reduce the expression of genes critical to myelopoiesis, leading to profound monocytopaenia and lymphocytopaenia. Therefore, we speculate that R398W, to some extent, is a *cis*-regulatory element-dependent LOF mutant (as demonstrated by transactivation assays) where its effect differs dramatically on different regulatory elements. In agreement with our previous study (**Chapter 3**), we also demonstrated that T354M, T355del and L359V displaying dominant negative activity in some settings, but not others. The mechanism(s), however, has not been fully resolved. We surmise that aberrant protein binding partnerships may confer the dominant negative activity or enhance transactivation activity.

In **Chapter 3**, we made use of WEMSA for the analysis of GATA2 binding to DNA. Specific DNA recognition and binding is a hallmark for transcription factors. Our WEMSA studies and protein modelling suggest that GATA2 missense mutations can have a significant effect on DNA binding. Ali *et al*. demonstrated at least 3 classes of GATA3 mutations based on EMSA studies [13]. Class 1 consists of GATA3 ZF2 mutations that result in a loss of DNA binding; Class 2 consists GATA3 ZF1 mutation that lead to reduced DNA binding and Class 3

represents GATA3 mutations that do not alter DNA binding or affinity but lose binding to a protein partner [13]. We observe a similar pattern in our WEMSA study (at least in the *TCRD* enhancer) and propose 3 classes for GATA2 ZF2 mutants, (1) enhanced DNA binding, (2) reduced DNA binding and (3) loss of DNA binding. For instance, L359V belongs to group 1 that displays enhance DNA binding affinity [1]. T354M, R362Q and R398W are group 2 mutants that have reduced DNA binding affinity. Although they are partial LOF mutants, luciferase reporter assays indicate that these mutants may still be able to recruit transcriptional machinery to the promoter and initiate transcription. Group 3 consists of T355del and Emberger syndrome mutants that do not bind to DNA, at least to the *TCRD* enhancer. Furthermore, the WEMSA result is reinforced by the luciferase reporter assays in which group 3 mutants display activity close to basal level (EV). Mutation in T355 and C373 alter conformation of ZF2, while R361L fails to recognize the WGATAR core motif.

The three-dimensional structure of the GATA2 would be a powerful tool to provide invaluable insight into the structural-functional relationship of the GATA2 mutations. Although the structure of the GATA2 protein had not been solved, we used the solved closely related murine GATA3:DNA-bound structure for comparison. Through homology protein modelling, we analysed the effect of GATA2 ZF2 missense mutation on protein-DNA interaction. Overall, the arginine residues comprise 12% of the ZF2 motif. Yet, arginine is one of the most frequently mutated residues along the ZF2 region (R361, R362, R396 and R398). DNA sequence specific recognition is a hallmark of all DNA binding TFs. This process occurs primarily in the major groove and often involves hydrogen bonding between TF amino acid side chains and nucleotides of the recognized motif [14]. In fact, several residues including threonine, leucine, arginine, asparagine have been shown to form hydrogen bonds and van der Waals contacts with DNA bases in the GATA3 ZF2 domain [9]. R361 is a key residue in the ZF2 domain that recognizes WGATAR DNA motif in the major groove by interlocking both sense and antisense base of the double stranded DNA. Other arginine residues such R362 and R398 probably serve as auxiliary amino acids that reinforce DNA binding and enhance or modulate specificity of the ZF2 domain to the DNA minor groove. Due to its unique physicochemical characteristics (positive charge), arginine is frequently found in protein active and DNA binding sites. In fact, arginine residues are often markedly enriched in the TF

binding sites, particularly in the domains that involve in the DNA binding in the minor groove [15]. Thus, arginine is a residue widely used by many TFs for specific protein-DNA recognition.

4.3.2 GATA2 Mutations Cause Aberrant Protein Partnerships

Aberrant protein-protein interactions such as enhanced interacting capacity have a significant impact on the cell homeostasis. In this study, we have provided evidence that GATA2 mutants (T354M and C373R) alter and enhance protein-protein interacting activity with PU.1. Our Co-IP assays indicated that T354M and C373R, at least in part, are GOF as shown by their enhanced capacity to bind to PU.1. Intriguingly, we did not see enhanced binding affinity for L359V when co-precipitated with PU.1 as claimed by Zhang *et al*. (2008) [2]. The inconsistency with their result is possibly due to the different cell types chosen for the experiments. These GATA2 mutant proteins may exert their GOF through a dominant interfering activity or by sequestering a putative co-factor(s) that interacts with GATA2. In support of our findings, Zhang *et al*. (1997) reported an N-terminally truncated RUNX1 may inhibit of RUNX1 activity by sequestering its co-activators [16]. We observed a dominant negative phenomenon in transactivation studies detailed in **Chapter 3**. We surmise that the dominant negative activity conferred by these GATA2 mutants is probably caused by the protein sequestration. The repression of PU.1 activity could, at least in part, be the result of these TFs being sequestered away from the transcriptional machinery by the GATA2 mutants since it has been reported that heterodimerization of GATA2 with PU.1, likely prevents both TFs from binding their consensus DNA elements [10, 17].

4.3.3 GATA2 Mutants Display Loss-of-Function in Haematopoietic Cells

Multiple lines of evidence have suggested that GATA2 operates in a critical dosage dependent manner, either maintaining the "stemness" and proliferation/survival of HSCs or promoting differentiation via down-regulation of its expression level and/or inhibiting its activity via protein-protein interaction during specification of cell fates [10, 18, 19]. Exogenous overexpression of GATA2 is a popular tool to study its function in haematopoietic cells. However, the results produced in these experiments are inconsistent and often contradictory each others. For instance, Persons *et al*. demonstrated that retrovirus mediated overexpression

of GATA2 WT blocks HPC derived colony formation [19]. Kitajima *et al*. reported prolonged activation (15 days) of GATA2 in ES cells promoted production of erythroid lineage cells [20]. Minegishi *et al*., on the other hand, described an indirect model in which a GFP reporter was inserted into the GATA2 locus of the transgenic mice. The GFP signal was not detected in the quiescent cells, but in the proliferating HPCs *in vitro* [21]. The inconsistency of the results is probably due to the cell types (embryonic cells vs HSCs vs HPCs), expression system (transient vs stable and constitutive vs inducible), GATA2 expression levels (high vs low) and duration (7 days vs 15 days) chosen for these experiments. In our experiments, we analysed the oncogenic potential of GATA2 mutants on murine whole BM using a stably transduced constitutively expressing retrovirus system, and assessed *in vitro* differentiation and proliferation in the presence of multiple myeloid differentiation factors. Our GATA2 WT results are in good agreement with the study by Persons *et al*. in which overexpression of GATA2 WT inhibits proliferation. Indeed, the high levels of GATA2 in HSC are thought to contribute to the relative quiescence of these cells [22, 23]. In the 2° assay, however, the total CFU increased by 2-fold. One possible explanation for the increased in total CFU is that prolonged exposure to multiple myeloid differentiation factors (*e.g*. G-CSF) may override the "stemness" effect of GATA2 via an unknown mechanism that drive the cells toward proliferation and differentiation pathways. In contrast, expression of constitutively overexpressed GATA2 mutants induced a pronounced expansion of haematopoietic progenitors. This result is consistent with our other functional data in which the mutants displayed LOF (or change-of-function). We speculate that GATA2 mutations may not as effectively provide a block in proliferation resulting in over active cycling of HSCs that can lead to leukaemia development or even depletion and exhaustion of the stem cell pool.

Total BM cells are a heterogeneous population that contains HSCs, HPCs and other supporting cells. Retrovirus can infect all of these cells. It is possible that the results generated by whole BM cells may mask the actual role of GATA2 in HSCs. For instance, the effect of GATA2 WT and mutants when introduced into late-committed progenitor cells might be different than when expressed in HSCs. Thus, the total BM cells may not be an ideal model to study a stem cell TF, such as GATA2. During the writing of this thesis, our research group has begun to isolate a more pure population of HSCs (LSK cells; Lin^- $Sca1^+$ Kit^+) from the BM. Our goal is

to decipher the specific and precise mechanism of these GATA2 mutants in leukaemia initiation and development. The combination of *in vitro*, *ex vivo*, and *in vivo* experimental approaches will provide new insights into the leukaemic mechanisms conferred by GATA2 mutants.

4.3.4 T354M is a Potent Mutation Affecting Myeloid Differentiation

Cell fate decisions are known to operate in a combinatorial action by groups of TFs (see **Figure 4.9a**). PU.1 is a master regulator of myeloid differentiation and interacts with many proteins via b3/b4 (a small region of ETS domain, see **Figure 4.9b**). Previous studies indicate that GATA2 maintains "stemness" of HSC via inhibition of PU.1 activity [10] and that this is lifted as cells differentiate and the mole ratio of GATA2 to PU.1 drops. Furthermore, studies have shown that GATA2 ZF2 domain is required for DNA binding and heterodimerization with the PU.1 protein [10]. This piece of information has prompted us to expect that T354M could also inhibit PU.1 function in a similar fashion. Hence, we propose that T354M manifests its oncogenic effects via both aberrant protein-protein interaction (GOF model) and by altering expression levels of critical GATA2 target genes (LOF model). The interaction between GATA1/2 and PU.1 is primarily through its ZF2 domain and the b3/b4 (part of the EST domain) of PU.1 [10, 24]. In the HSC, WT GATA2 prevents differentiation by blocking the c-JUN and PU.1 interaction [10] (see **Figure 4.10a**). Unlike WT GATA2, T354M has a stronger affinity for PU.1. In the proposed GOF model, heterodimerization of T354M and PU.1 prevents dissociation of PU.1 from the complex, thus, blocking expression of PU.1 target genes such as *M-CSF* [25], or it sequesters and prevents it PU.1 from interacting with its co-activator, c-JUN (see **Figure 4.10b**). Thereby, T354M prevents HSC differentiation by inactivating the myeloid gene expression program.

Previous studies have demonstrated that GATA2 dosage is important in regulation of HSC production, survival and expansion [18, 26]. T354M has a weaker DNA binding affinity resulting in altered expression of GATA2 target genes [1]. Isothermal titration calorimetry showed that T354M has approximately half (56.3%) the DNA binding activity of WT (see **Chapter 3**). One might propose then that in heterozygous T354M mutant cells, there is only ≈

75% of DNA binding activity. Thus, T354M fits into the proposed LOF model in this context (**Figure 4.10c**).

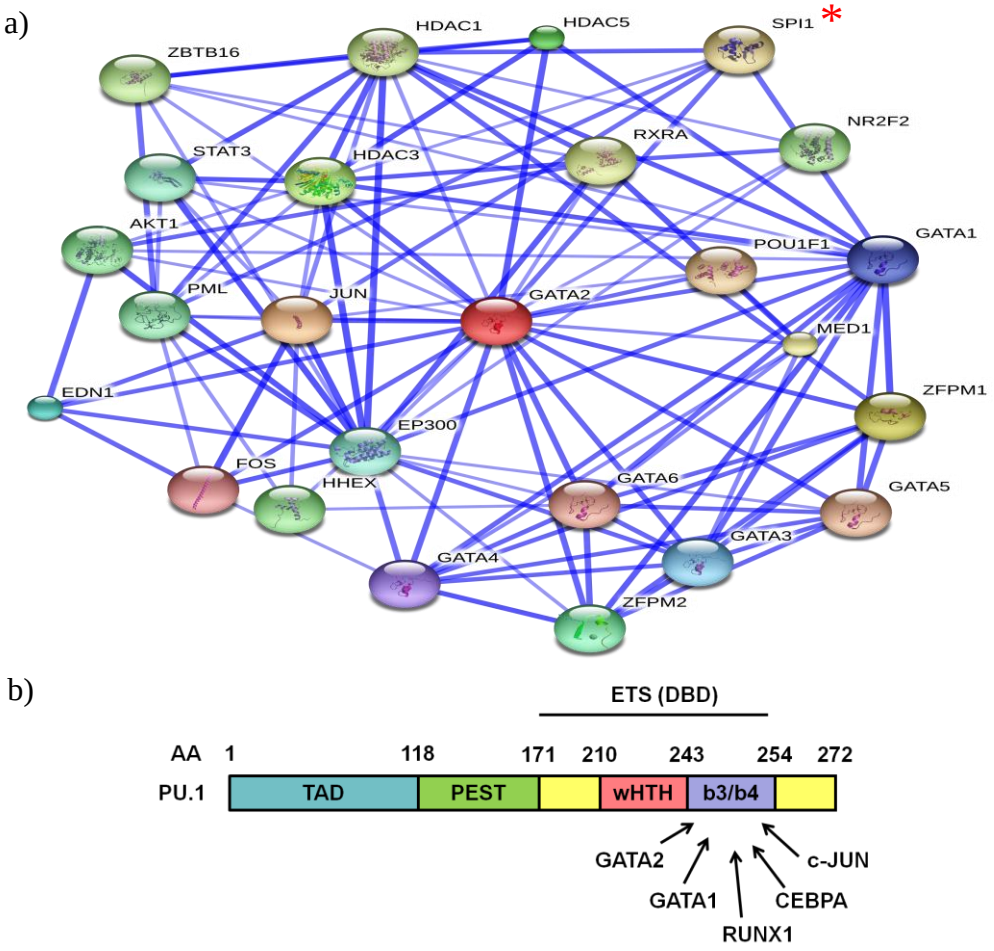


Figure 4.9: Participation of GATA2 in the Haematopoietic Transcriptional Regulatory Network.

a) **The haematopoietic transcriptional regulatory network.** GATA2 interacts with many proteins, such as the GATA family members and SPI1* (PU.1). Many of these proteins are also involved in haematological malignancies. Stronger protein-protein interactions or associations are represented by thicker lines. The transcriptional regulatory network was generated using STRING 9.05 [27]. b) **PU.1 functional domains.** A relatively small region of the PU.1 ETS domain (b3/b4) mediates inhibitory interactions with GATA1/2, as well as cooperative and/or synergistic positive interactions with RUNX1 and CEBPA. c-JUN acts an important coactivator of PU.1 function. Abbreviations: AA, amino acid; TAD, transactivation domain; PEST, proline (P), glutamic acid (E), serine (S), and threonine (T) riched domain; ETS, Eleven-twenty six domain; DBD, DNA binding domain; wHTH, winged helix-turn-helix domain. (Picture adapted and redrawn based on Zhang *et al* [10]).

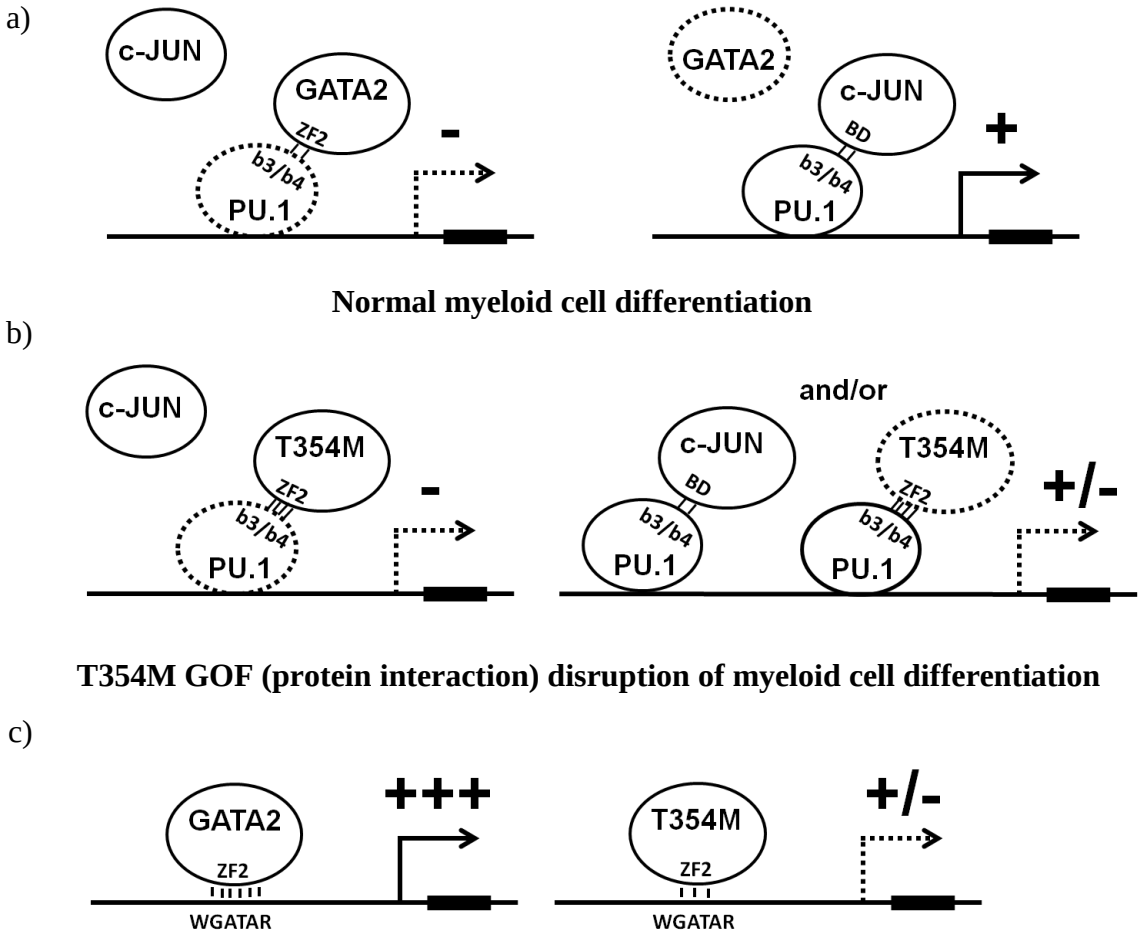


Figure 4.10: Gain-of-Function and Loss-of-Function Models for GATA2 Mutation.

a) **Normal differentiation of the myeloid cells.** Myeloid cell differentiation through protein-protein interactions proposed by Zhang *et al* [10]. In the HSC, high GATA2 levels block transcription of PU.1 downstream target genes by competing with c-JUN for interaction with PU.1 [10], thus, preventing PU.1:c-JUN interaction (left). During myeloid differentiation, GATA2 expression is reduced and released from the complex, hence, PU.1 can work synergistically with c-JUN to transactivate PU.1 target genes (right) [10]. b) **GOF model for T354M.** T354M has an enhanced interacting activity. Heterodimerization of T354M and PU.1 may prevent PU.1:c-JUN interaction and blocks PU.1 target genes such as *M-CSF* (left) [25]. As GATA2 levels decrease during differantioan, T354M remains more tightly bound to PU.1, hence sequestering PU.1 from interacting with its co-activator, c-JUN (right). c) **LOF model for T354M.** GATA2 WT recognizes and binds strongly to WGATAR consensus sequence activating its target genes (left). T354M has a weaker DNA binding affinity compared to WT, thereby, reducing the expression of GATA2 target genes (right). Abbreviations: ETS, Eleven-twenty six domain; BD, Basic domain; ZF2, Zinc Finger 2 domain; WGATAR, consensus DNA binding motif; Dotted lines, low expression.

4.4 Conclusion

We conclude that GATA2 conserved residues cannot be replaced by other residues and retain appropriate GATA2 function. Each of the mutations is such that they confer specific but subtle changes to GATA2 function, which ultimately leads to a disease phenotype for immunodeficiency and lymphoedema or to a selective advantage for MDS/AML. Substantial additional work is still required to fully understand their function in haematological malignancy development.

4.5 Recommendations for Future Work

Considerable progress has been made on the front of identifying *GATA2* mutation as a predisposition gene for a range of diseases. Although *in vitro* functional assays presented in this chapter have given us a glimpse into the possible mechanisms conferred by these mutants, we are still far from understanding its molecular basis. How *GATA2* mutation can cause three distinct diseases (MDS/AML, immunodeficiency and lymphoedema) *in vivo* is still remains an enigma wrapped in a riddle. Recognition that *GATA2* is an important disease causing gene, raises two fundamental questions which should be addressed in the future studies:

1. **Is the disease-causing mechanism associated with aberrant protein-protein interaction?** In this chapter, we have demonstrated that 2 GATA2 mutants (T354M and C373R) have an altered interacting activity with their co-factors. Our findings have shed some light on the possible function on how these mutants operate *in vivo*. Other than affecting DNA binding, we recognize that an enhanced or loss of interacting activity with either co-activators or co-repressors could be an important mechanism by which *GATA2* mutants initiate and drive disease progression. However, our Co-IP experiment is limited to detect known GATA2 interacting proteins. During the writing of this thesis, our research group has begun to investigate interaction of GATA2 with putative endogenously expressed co-factors in a more relevant myeloid cell lines, HL-60 cells, using a cutting edge technology known as stable isotope labelling by amino acids in culture (SILAC)-based quantitative mass-spectrometry.

This approach allows us to perform an unbiased genome-wide comparative proteomic analysis for GATA2 binding partners on HL-60 cells expressing GATA WT or mutants. This technique can identify novel GATA2 protein binding partners that interact with particular mutants and quantify any changes in binding ability.

2. **How do *GATA2* mutants fit into a disease model in the haematopoietic system?** An important question that my study was not able to address is how and to what extent *GATA2* is involved in haematological disorders *in vivo*. In **Chapter 3**, we have addressed that GATA2 mutants have a significant effect on proliferation, differentiation and apoptosis in HL-60 cells. HL-60 cells are an invaluable model for studying cellular oncogene expression in relation to particular hematopoietic differentiative lineages [28]. However, HL-60 cells are a leukaemic cell line with cancerous properties including karyotypic abnormality, unlimited proliferative capacity and tumour formation in an animal model [29]. To accurately address GATA2 function in haematopoiesis, we have begun to isolate mouse HSCs (LSK cells) and use them for *ex vivo* genetic manipulation. These genetically manipulated adult stem cells serve as a powerful tool from which to address many of the key biological processes in the haematopoietic system. For instance, *in vitro* tissue culture assays on these cells will give us an important clue regarding the clonogenic and differentiation capacity conferred by the mutants. GATA2 (WT or mutants) expression in these retrovirally transduced LSK cells will be checked throughout the serial replating experiment. Ultimately, these modified LSK cells will be transplanted back into congenic recipient mice. It would be very interesting to know whether GATA2 mutants alone can cause disease (*e.g.* immunodeficiency syndrome and haematological malignancies). How aggressive are the diseases? Are these mice susceptible to infection as seen in the patients with immunodeficiency syndrome? Do these mice need to be stressed to promote disease? Investigations into the aforementioned issues are warranted.

4.6 References

1. Hahn, C.N., et al., *Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia.* Nature genetics, 2011. **43**(10): p. 1012-7.
2. Zhang, S.J., et al., *Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia.* Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2076-81.
3. Ostergaard, P., et al., *Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome).* Nature genetics, 2011. **43**(10): p. 929-31.
4. Yan, X.J., et al., *Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia.* Nature genetics, 2011. **43**(4): p. 309-15.
5. Dickinson, R.E., et al., *Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.* Blood, 2011. **118**(10): p. 2656-8.
6. Hsu, A.P., et al., *Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome.* Blood, 2011. **118**(10): p. 2653-5.
7. Ko, L.J., et al., *Murine and human T-lymphocyte GATA-3 factors mediate transcription through a cis-regulatory element within the human T-cell receptor delta gene enhancer.* Mol Cell Biol, 1991. **11**(5): p. 2778-84.
8. Bert, A.G., et al., *A modular enhancer is differentially regulated by GATA and NFAT elements that direct different tissue-specific patterns of nucleosome positioning and inducible chromatin remodeling.* Mol Cell Biol, 2007. **27**(8): p. 2870-85.
9. Bates, D.L., et al., *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, 2008. **381**(5): p. 1292-306.
10. Zhang, P., et al., *Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8705-10.
11. Rodrigues, N.P., et al., *GATA-2 regulates granulocyte-macrophage progenitor cell function.* Blood, 2008. **112**(13): p. 4862-73.
12. Lavau, C., et al., *Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL.* The EMBO journal, 1997. **16**(14): p. 4226-37.
13. Ali, A., et al., *Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor.* Hum Mol Genet, 2007. **16**(3): p. 265-75.
14. Garvie, C.W. and C. Wolberger, *Recognition of specific DNA sequences.* Mol Cell, 2001. **8**(5): p. 937-46.
15. Rohs, R., et al., *The role of DNA shape in protein-DNA recognition.* Nature, 2009. **461**(7268): p. 1248-53.

16. Zhang, Y.W., et al., *A novel transcript encoding an N-terminally truncated AML1/PEBP2 alphaB protein interferes with transactivation and blocks granulocytic differentiation of 32Dcl3 myeloid cells.* Mol Cell Biol, 1997. **17**(7): p. 4133-45.
17. Walsh, J.C., et al., *Cooperative and antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates.* Immunity, 2002. **17**(5): p. 665-76.
18. Ling, K.W., et al., *GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells.* The Journal of experimental medicine, 2004. **200**(7): p. 871-82.
19. Persons, D.A., et al., *Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis.* Blood, 1999. **93**(2): p. 488-99.
20. Kitajima, K., et al., *GATA-2 and GATA-2/ER display opposing activities in the development and differentiation of blood progenitors.* The EMBO journal, 2002. **21**(12): p. 3060-9.
21. Minegishi, N., et al., *Expression and domain-specific function of GATA-2 during differentiation of the hematopoietic precursor cells in midgestation mouse embryos.* Blood, 2003. **102**(3): p. 896-905.
22. Ezoe, S., et al., *GATA-2/estrogen receptor chimera regulates cytokine-dependent growth of hematopoietic cells through accumulation of p21(WAF1) and p27(Kip1) proteins.* Blood, 2002. **100**(10): p. 3512-20.
23. Tipping, A.J., et al., *High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle.* Blood, 2009. **113**(12): p. 2661-72.
24. Nerlov, C., et al., *GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription.* Blood, 2000. **95**(8): p. 2543-51.
25. Behre, G., et al., *c-Jun is a JNK-independent coactivator of the PU.1 transcription factor.* J Biol Chem, 1999. **274**(8): p. 4939-46.
26. Lim, K.C., et al., *Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning.* J Clin Invest, 2012. **122**(10): p. 3705-17.
27. Szklarczyk, D., et al., *The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.* Nucleic acids research, 2011. **39**(Database issue): p. D561-8.
28. Collins, S.J., *The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression.* Blood, 1987. **70**(5): p. 1233-44.
29. Gallagher, R., et al., *Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia.* Blood, 1979. **54**(3): p. 713-33.

Chapter 5: Effect of GATA2 Mutations in the Lymphatic System

5.1 Introduction

The human lymphatic vascular system plays essential physiological functions by maintaining body fluid homeostasis, nutrient absorption and immune cell trafficking. Due to its crucial and indispensable roles in normal physiology, it is perhaps not surprising that lymphatic vasculature dysfunction has been implicated in a variety of human diseases, including lymphoedema and a compromised immune system.

Hereditary lymphoedema syndrome is a rare genetic, autosomal disease that results from a malfunctioning lymphatic vascular system, leading to a chronic and debilitating condition caused by progressive accumulation of interstitial fluid and formation of nonpitting localized tissue swelling [1]. The disease generally manifests years or decades after birth, but can develop *in utero* and in newborn infants [1]. Several genes that appear to selectively influence lymphatic growth and remodelling have been associated with familial lymphoedema development. Karkkainen *et al*. (2000) first reported the lymphoedema-linked missense mutations of *VEGFR3* in 4 families with Milroy disease, a disease caused by underdeveloped and dysfunctioning cutaneous lymphatic vessels [2]. Functional assays indicated that these mutant proteins have prolonged cellular half-lives and lack of tyrosine-kinase activity [2]. In the same year, *FOXC2* mutations (a nonsense mutation and a frameshift mutation) were reported by Fang *et al*. in 2 families with the lymphoedema-distichiasis syndrome [3]. These patients present with primarily lymphoedema with variable age at onset, together with distichiasis (double rows of eyelashes). Genetics underlying hypotrichosis-lymphoedema-telangiectasia syndrome (HLT) was deciphered by Irrthum *et al*. [4]. They reported that the dominant and recessive inheritance of *SOX18* mutations are associated with HLT, highlighting that *SOX18* is essential for the development and/or maintenance of lymphatic vessels [4]. Alders and co-workers discovered that mutations in *CCBE1* cause Hennekam syndrome, an

autosomal recessive disease in humans characterised by intestinal lymphangiectasia (a rare disorder characterized by dilated and functionally obstructed intestinal and mesenteric lymphatic vessels [5]) and lymphoedema, with facial anomalies and severe mental retardation [6, 7].

We and others have established that *GATA2* mutations are associated with primary lymphoedema and Emberger syndrome [8-10]. GATA2 is an important transcription factor that is required for both haematopoiesis and lymphatic development. Given that GATA2 is abundantly expressed in the human vein endothelial cells [11], it is not surprising that GATA2 deficiency can cause vascular defects. Moreover, several developmentally regulated endothelial-specific enhancers and promoters including *Cdh5* [12], *Epcr* [13], *Flk1/Kdr* (encoding VEGFR2) [14], *PECAM1* [15], *ICAM2* [16] and *Lyl1* [17] have been shown to be bound directly by GATA2. Further, in a *Gata2*-GFP knockin mouse, GFP expression was detected in the endocardial, vascular and lymphatic endothelium, particularly in the budding lymphatic endothelial cells (LECs) and the lymphatic vessels after birth [18]. Moreover, recent work has revealed that conditional genetic ablation of *Gata2* (*VE-Cre*) results in oedema and haemorrhage in mouse embryos caused by defective lymphatic development [19].

In our previous study, we demonstrated that GATA2 is highly expressed in the embryonic and adult mouse lymphatic vessels, particularly in the leaflets of lymphatic vascular valves [8]. Intriguingly, expression of several lymphatic valve development and maintenance genes in primary LECs including *Prox1*, *Foxc2*, *Angpt2* and *Itga9* were significantly down-regulated upon knockdown of *Gata2* [8]. This result has prompted us to investigate the connection between *Gata2* and these genes. In a search for novel TFs that can regulate *Prox1* expression, we screened and identified putative GATA binding sites in the intronic DNA of the *Prox1* gene. We also demonstrated that GATA2 can bind to and transactivate gene expression via these GATA binding sites, suggesting that GATA2 is a potential regulator of *Prox1*.

5.2 Results

5.2.1 GATA2 Transactivates the *Prox1* Promoter/Enhancer Element

In our previous study, we demonstrated that high levels of GATA2 and PROX1 are co-expressed in the lymphatic vessel valves. Furthermore, we also demonstrated that knockdown of *Gata2* mRNA leads to down-regulation of genes important for programming lymphatic valve development including *Prox1* [8]. We therefore asked whether GATA2 could regulate *Prox1* expression. We first searched for GATA2 binding sites within the mouse *Prox1* gene (transcribed to NM_008937) and surrounding regions using published ChIP-Seq data [20]. This highlighted 2 regions: a 4kb region in intron 1 and a region at -11.3kb from the *Prox1* promoter 1 (P1) (**Figure 5.1a**). Both regions contained putative consensus GATA binding sites. Interestingly, Francois *et al*. identified an alternative promoter (P2) immediately upstream of exon 2 encompassing the GATA binding sites which was shown to bind and transcribe in response to SOX18 [21]. Hence, we cloned this murine *Prox1* P2 DNA fragment (4kb; we call this *Prox1* promoter/enhancer or *PEE*) into a luciferase reporter plasmid. The *PEE* region contained 5 putative WGATAR consensus sequences (**Figure 5.1a**). A putative *Prox1* enhancer that was also identified via GATA2 ChIP-Seq data from the murine haematopoietic progenitors cell line, HPC-7 [20], is located approximate 11.3kb upstream of the *Prox1* transcription start site (P1) (designated *Prox1* -11.3kb; **Figure 5.1a**). To explore the role of *Gata2* in lymphatic vascular development, we cloned the *PEE* region into a luciferase report construct and analyzed the effect of GATA2 (WT or mutants) expression on luciferase activity (**Figure 5.1b**). Overexpression of WT GATA2 led to a significant increase in luciferase activity, implicating functional relevance of GATA2 on *PEE* (**Figure 5.1c**). Intriguingly, all GATA2 mutants except for L359V significantly reduced transactivation of *PEE*. Hence, GATA2 can regulate *Prox1* expression, at least at the *PEE* level, and a reduction in *Prox1* expression *in vivo* could be expected with the GATA2 mutants.

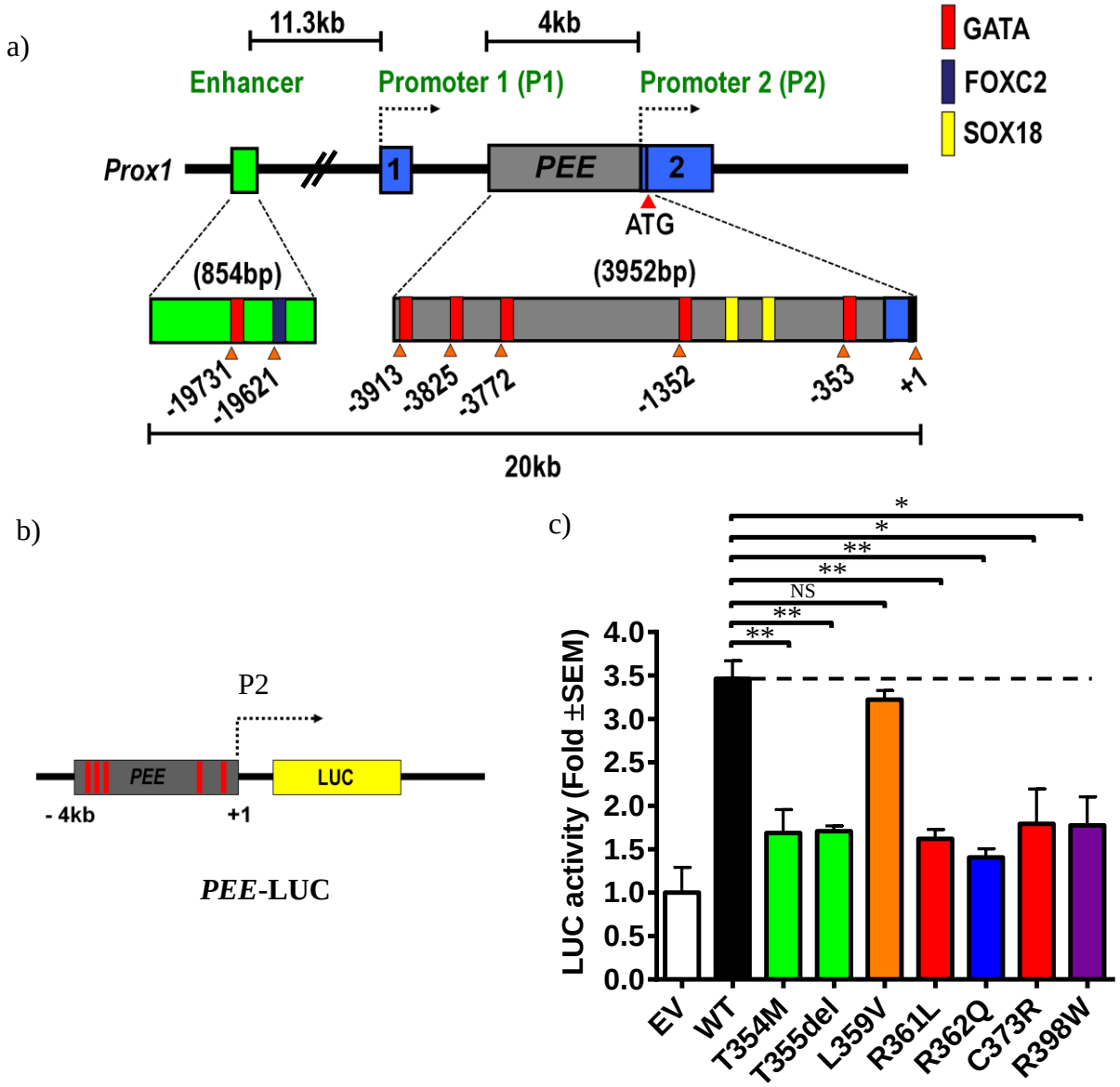


Figure 5.1: *Prox1* Enhancer/Promoter Elements Contain Functional GATA Sites.

a) **Genomic organization of the murine *Prox1* gene.** *Prox1* is transcribed from the P1 where the start codon (ATG) is located in the exon 2 (RefSeq: NM_008937). The *Prox1 PEE* is an intronic region which encompasses the immediate 4kb upstream from the exon 2 and has been shown to have promoter activity [21]. Computational analysis revealed 5 putative GATA sites in the *PEE* region, which are located at the position of -3913bp, -3815bp, -3772bp, -1352bp and -353bp upstream of the (+1) ATG transcription initiation site. We also identified a putative *Prox1* enhancer region at 11.3kb (green) upstream of the P1. A putative GATA site (-19,731bp from +1) and a putative FOCXC2 site (-19,621bp from +1) were identified in the enhancer. b) **The *PEE* luciferase reporter construct.** The *PEE* was sub-cloned into a luciferase reporter construct. c) **Transactivation of GATA2 (WT or mutants) on the *PEE* luciferase reporter construct.** The *PEE*-LUC construct was cotransfected into HEK293 cells with GATA2 WT or mutants and luciferase activity was measured. Luciferase activity is shown as fold change compared to EV and differences between transactivation of WT and mutants was determined. $^{**}P<0.01$, $^{*}P<0.05$ and NS: not significant.

5.2.2 GATA2 Binds to *Prox1* Promoter/Enhancer Element

To further investigate the possible role of *Gata2* in modulating *Prox1* expression, we established DNA binding assays on these putative GATA sites. Short oligonucleotide probes corresponding to each of the binding sites were synthesized and subjected to WEMSA (see **Appendix C-2, Supplementary Table 4** for probes). These studies validated GATA2 binding to *PEE*, particularly to the GATA sites at the position of -1352 and -353 (**Figure 5.2a**). Weak binding was also detected to the -3913 and -3825, but not in the -3372 putative GATA sites. The *PEE* was then reduced from 4kb to 1.5kb, removing the 3 weak binding putative GATA sites, but maintaining the strong GATA2 binding site (we called this region sub-*PEE*) (**Figure 5.2b**, left panel). The -1352 and -353 GATA sites were then mutagenised from GATA to CTTA individually and together to destroy the GATA2 binding sites. These sub-*PEE* DNA fragments were then sub-cloned into a luciferase reporter construct, and were then called s*PEE*-LUC (WT GATA binding sites), s*PEE*-m1-LUC (mutated -1352 GATA site), s*PEE*-m2-LUC (mutated -353 GATA site) and s*PEE*-m1m2-LUC (mutated -1352 and -353 GATA sites). Overexpression of GATA2 increased luciferase activity when using the s*PEE*-LUC construct that contains both GATA binding sites (**Figure 5.2b**, right panel). The specificity of GATA2 driven luciferase activity on sub-*PEE* constructs was confirmed using the s*PEE*-m1-LUC, s*PEE*-m2-LUC and s*PEE*-m1m2-LUC constructs. Of note, mutation of WGATAR sequence in -1352 significantly reduced luciferase activity induced by GATA2. However, we observed no significant effect on the singly mutated -353 site. Intriguingly, alteration of both the -1352 and -353 GATA2 binding sites almost completely abolished the transactivation activity by GATA2. Together, the WEMSA and reporter assays suggest that -1352 is an important *cis*-acting element for *Prox1* expression and that the -353 binding site also contributes in a significant way.

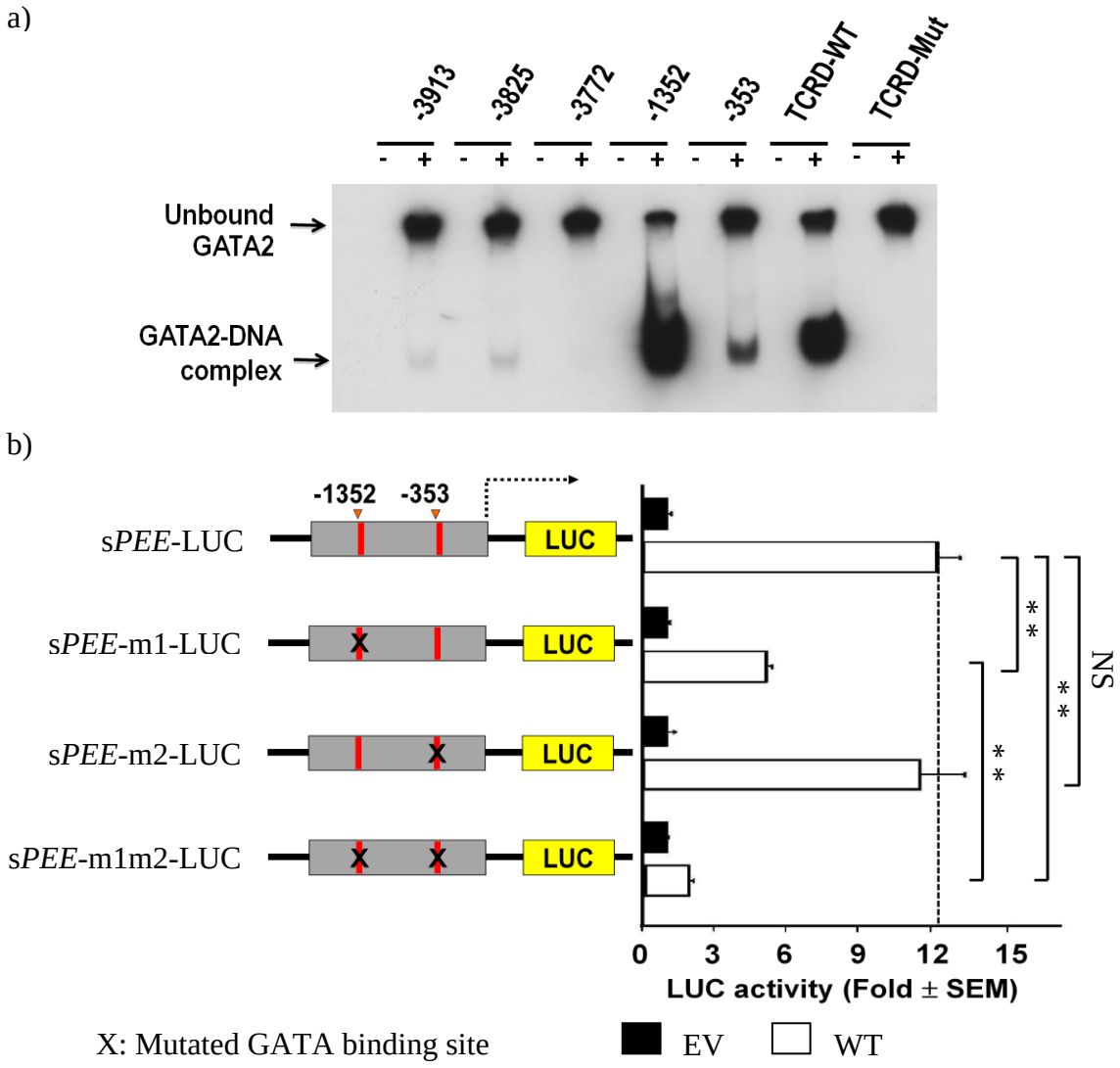


Figure 5.2: Functional Characterization of *PEE*.

a) **WEMSA on *PEE*.** Oligonucleotides corresponding to each GATA binding site were incubated either with EV (-) lysate or GATA2 WT (+) lysate. Human *TCRD* enhancer WT and mutant probes were used as positive control (WT GATA binding site) and negative control (mutated GATA binding site), respectively. b) **Transactivation assays on *PEE*.** A *PEE* sub-fragment (-1.5kb to +1bp) containing 2 putative GATA binding sites, -1352 and -353, was sub-cloned into a luciferase reporter construct (Top left panel). To mutate the GATA2 binding site(s), the GATA was changed to CTTA (indicated by the X). A *Gata2* WT or EV expression construct was cotransfected with either s*PEE*-LUC, s*PEE*-m1-LUC, s*PEE*-m2-LUC or s*PEE*-m1m2-LUC into HEK293 cells. Luciferase activity for each of the combinations was normalized to the respective empty vector control (right panel). ***P*<0.01 and NS: not significant.

5.2.3 GATA2 WT and Mutants Exhibit Differential Binding Affinity to GATA2-Responsive Elements in *PEE* and Human Haematopoietic Lineage Promoters

To further dissect the roles of *GATA2* mutation and its association with different phenotypes, we performed DNA binding assays (WEMSA) on several known GATA2 responsive promoter/enhancer elements (**Figure 5.3**). In **Chapter 4**, we demonstrated that GATA2 mutants (human) bind to human *TCRD* enhancer to varying degrees. For DNA binding studies to the mouse *Prox1* promoter/enhancer GATA2 binding sites, we chose also to generate the whole GATA2 mutant allelic series for the mouse GATA2 protein. When these were used in WEMSA on the human *GM-CSF* promoter and *TCRD* enhancer elements (**Figure 5.3a** and **b**), we saw the same DNA binding affinity pattern as for the corresponding human GATA2 mutants. Hence, we saw excellent concordance for human and mouse GATA2 (WT and mutants) proteins on the haematopoietic lineage target DNA (see **Chapter 4; Figure 4.3**).

We then asked if the *PEE* is bound by GATA2 WT and mutants. WEMSA studies revealed a significantly different binding capacity to -1352 and -353 probes by GATA2 proteins. In general, DNA binding affinity to the -1352 probe for mouse GATA2 proteins is similar to the DNA binding pattern seen in the haematopoietic lineage probes (**Figure 5.3c**). However, T354M, R362Q and R398W seemed to bind to the -1352 probe at levels close to or slightly less than WT and L359V. In contrast, T355del and R361L bound weakly to target DNA, and no DNA binding was seen in C373R. WEMSA on the -353 oligo probe, however, demonstrated a different DNA binding spectrum for these mutants. T354M, R362Q and R398W had a weaker DNA binding affinity, followed by T355del, L359V and R361L (**Figure 5.3d**). We did not detect DNA binding activity mediated for C373R. Taken together, we summarized the DNA binding pattern for the GATA2 responsive elements and broadly divided them into 3 major groups. Group 1 belongs to the mutants with variable binding affinity to DNA. L359V is a typical Group 1 mutant that displays both enhance (*TCRD* and *GM-CSF*) and reduced (*Prox1* -353) DNA binding affinity compared to WT. Group 2 mutants generally exhibit weaker DNA binding than WT, but still retain some binding activity. T354M, R362Q and R398W belong to Group 2. Group 3 mutants such as T355del, R361L and C373R did not bind or bind poorly to DNA. Our classification not only highlights the discrepancy for

DNA binding among the GATA2 mutant groups, but may also reflect their contributions to the development of specific disease. For example, Group 2 mutants often have reduced DNA binding and are associated with MDS/AML and immunodeficiency syndrome. In contrast, complete LOF function mutants (except T355del) are associated with MDS/AML, immunodeficiency syndrome and lymphoedema.

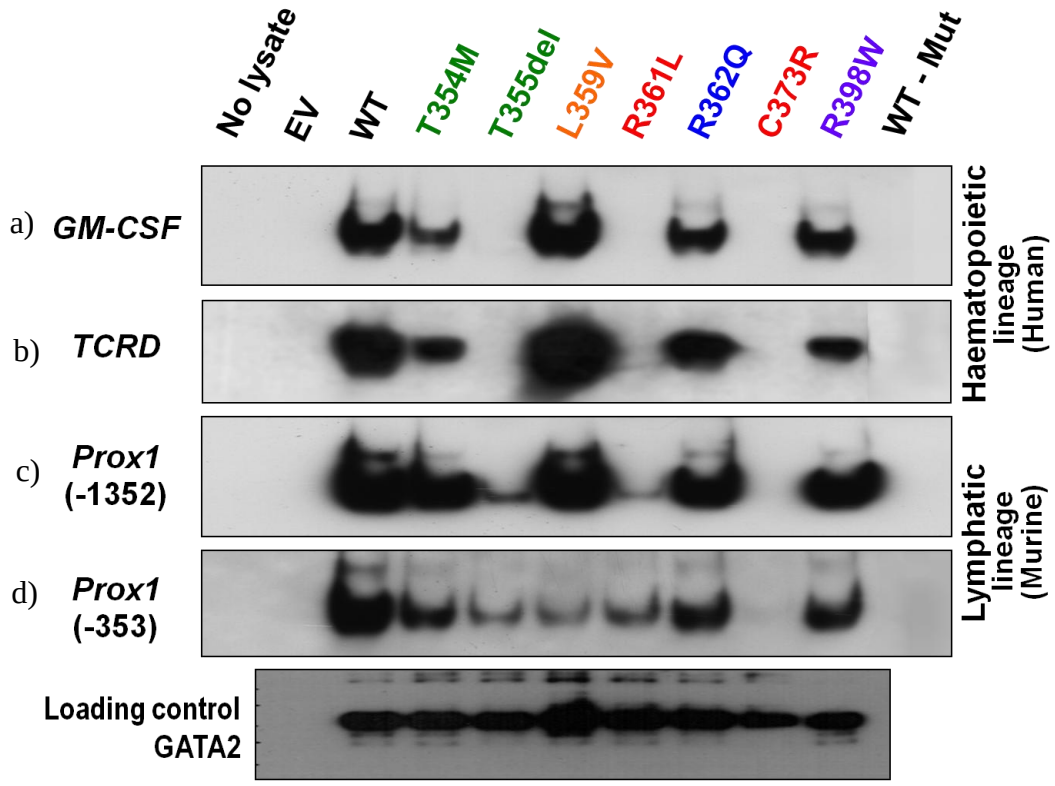


Figure 5.3: WEMSA for GATA2 WT and Mutants on a Panel of Known Responsive Promoter/Enhancer Elements.

A comparable amount of GATA2 nuclear lysate for WT or mutants (mouse) was incubated with short oligonucleotide probes corresponding to promoter/enhancer elements from human haematopoietic lineages (a) and (b), and murine *Prox1 PEE* (c) and (d). WEMSA blots were probed with GATA2 antibody.

5.2.4 *Prox1* -11.3kb Enhancer is Transactivated by Multiple Transcription Factors

The *Prox1* gene -11.3kb enhancer is conserved between mice and humans (91.5% identical). ChIP-Seq data from ENCODE [22] (via UCSC Genome Browser [23]) shows an active

enhancer signature such as H3K27Ac, is found in the human counterpart (HUVEC cells) (**Figure 5.4a**, top panel). Importantly, multiple TFs including GATA2 have been shown to bind to this region. We reasoned that the -11.3kb enhancer is a long-range regulatory element for *Prox1*, and an 854bp DNA fragment of this enhancer, that encompasses the GATA and FOXC2 binding site, was cloned into a luciferase reporter construct (*Prox1*-Enh-LUC) (**Figure 5.4a**, bottom panel). Overexpression of GATA2 WT resulted in a 2.5-fold increase in luciferase activity, demonstrating that GATA2 can transactivate this enhancer. However, a significant reduction in luciferase activity was seen for all mutants except the CML-BC L359V mutant (**Figure 5.4b**, top panel). WEMSA of the -11.3kb enhancer sole putative GATA binding site showed variable binding affinity of GATA2 mutants similar to that observed for the -353 probe except for very weak DNA binding for T355del (**Figure 5.4b**, bottom panel). Note that, the level of DNA binding to this element does not closely correspond to the transactivation levels of the entire -11.3kb enhancer, suggesting that other TFs also play a significant role.

We set out to dissect the function of the enhancer by cotransfecting the -11.3kb luciferase construct together with several lymphatic specific TF expression plasmids. In the individual TF assays, GATA2, FOCX2 and SOX18 exhibited enhanced luciferase activity above EV, while PROX1 inhibited reporter activity (**Figure 5.4c I**). Transcription factors are known to form a multimeric transcription complexes at important transcriptional elements. To explore the role of GATA2 and its protein partnerships in *Prox1* transcription initiation, several known TFs that are required for lymphatic development were cotransfected into HEK293 cells together with GATA2. In the dual TF assays, the coexpression of PROX1 with GATA2 reduced the effect of GATA2 transactivation from 2.0 fold to 1.5 fold, consistent with PROX1 inhibition of this enhancer seen for PROX1 alone (**Figure 5.4c II**). In contrast, GATA2/FOXC2 and GATA2/SOX18 increased reporter activity 9.6- and 5.6-fold, respectively, suggesting that these TFs may work synergistically to activate this enhancer. Hence, our experiment data suggests that GATA2 may form works together with other lymphatic TFs in regulating *Prox1* expression. Importantly, PROX1 itself may negatively feedback to regulate its own expression.

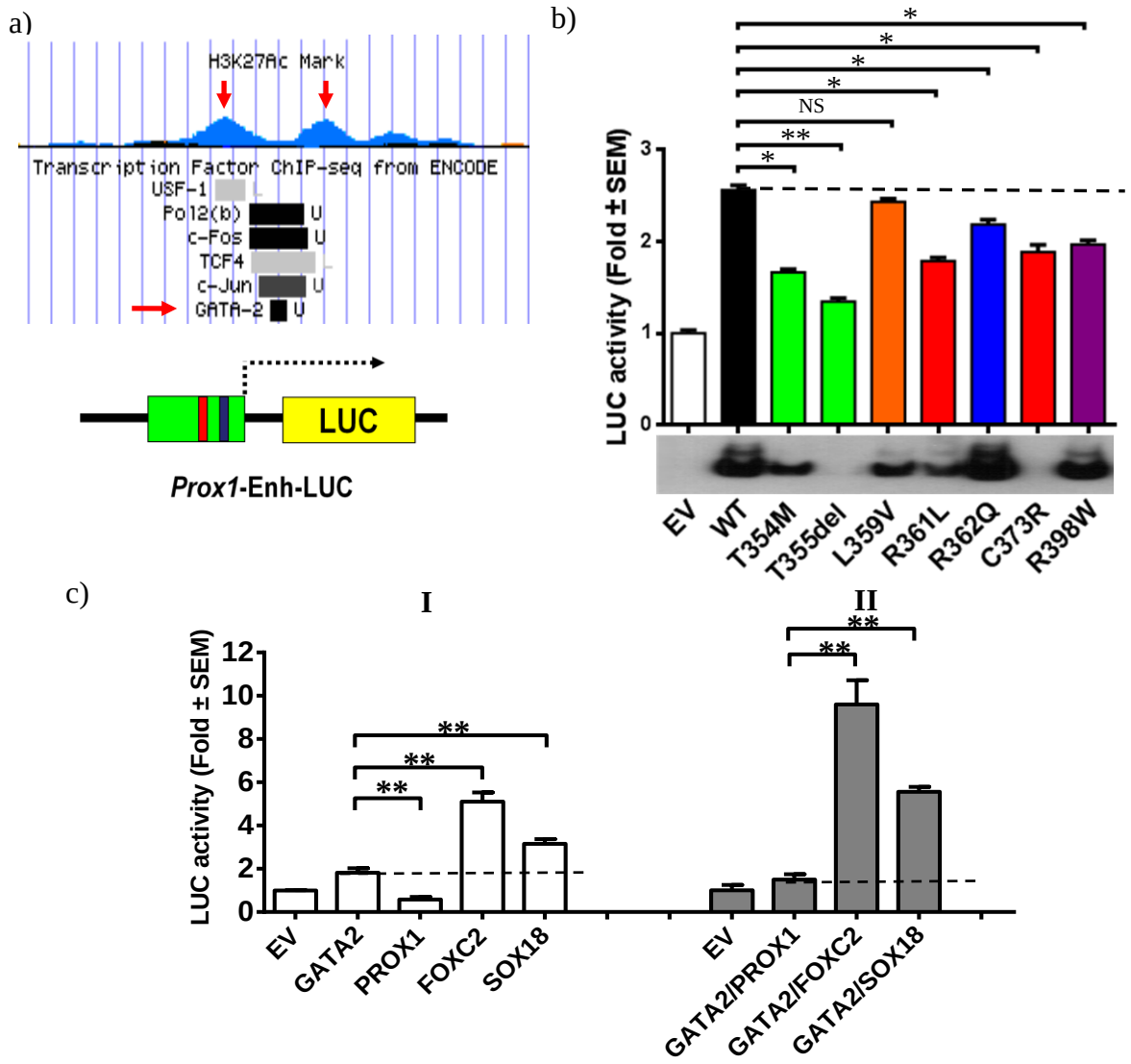


Figure 5.4: WEMSA and Luciferase Reporter Assays on the *Prox1* -11.3kb Enhancer.

a) **ChIP-Seq data on the putative human *PROX1* enhancer.** The data shows that multiple TFs bind to the -11.3kb region (top panel). The -11.3kb (green) murine *Prox1* DNA fragment was cloned into a luciferase reporter construct (bottom panel). The *Prox1*-Enh-LUC construct was then used in the transactivation studies in b) and c). b) **Luciferase reporter assays with the GATA2 mutant allelic series.** All mutants exhibited a significant reduction in luciferase activity except for L359V (top panel). **WEMSA study on murine *Prox1* -11.3kb enhancer.** GATA2 mutants exhibited differential DNA binding affinity to the oligo probe corresponding to the *Prox1*-11.3kb enhancer (bottom panel). No GATA2 WT binding was seen when the GATA binding site is mutated (data not shown). c) ***Prox1*-Enh-LUC construct transactivation by multiple TFs**. The *Prox1*-Enh-LUC construct was cotransfected with a *Gata2* WT or other lymphatic TFs expression construct. (I). The *Prox1*-Enh-LUC construct was cotransfected with a *Gata2* WT and one of the lymphatic TF constructs (II). Luciferase activity for each of the assays was normalized to the respective empty vector control. ***P*<0.01, **P*<0.05 and NS: not significant.

5.3 Discussion

5.3.1 GATA2 is Regulator for *Prox1* Expression

Over the past decade, there have been copious studies to decipher the molecular processes underlying the development of the lymphatic system. *VEGFR3*, *SOX18*, *FOXC2* and *PROX1* are several known genes that play important role in the lymphatic development. Homozygous knockout of these genes, except for *Sox18* [24], is embryonic lethal, a result caused by severe defects in cardio- or lymphatic vasculature [25]. Familial lymphoedema syndromes are known to have a strongly penetrant genetic component. Yet, only a handful of genes have been implicated in the development of the disease. For instance, germline mutations in *SOX18* and *FOXC2* are strongly associated with lymphoedema in humans [3, 4]. In the present study, we investigated the correlation of the lymphatic master regulator, *Prox1* expression with that of GATA2 expression and mutation. We looked into two previously reported intronic regions for the signs of enhancer/promoter signature. The first region, *Prox1* 4kb was reported by Francois *et al* [21]. They reported evidence that an alternative *Prox1* promoter (4kb within intron 1) is activated by SOX18 *in vitro* and *in vivo* [21]. Computational analysis indicated that there are 5 putative GATA sites in the 4kb *PEE* fragment, two of which (-1352 and -353) responded strongly to DNA binding and reporter gene assays. We then data mined for ChIP-Seq results available in the public domains and cross checked with data from the model cell lines. However, there are limited ChIP-Seq data for the *Prox1* expressing cells in the databases. To our best knowledge, there is no evidence linking *PROX1* expression to the model cell lines such as HUVEC (human umbilical vein endothelial cells) and K562 cells. Nevertheless, we speculate that GATA2 acts at the *Prox*1 locus as a "pioneer" TF. Our statement is supported by the studies that indicated HNF3 (also known as FOXA) and GATA4 occupy regulatory sites in chromatin prior to activation of albumin gene expression in pluripotent gut endoderm [26, 27]. Pre-occupancy of pioneer TF on silent chromatin endows transcriptional competence to the cells with rapid response to transcriptional signals [27], stabilizing nucleosomal structure for the binding of other TFs [28] or opening of the compacted chromatin for other TFs to bind [29]. GATA2 may poise on the *Prox1* enhancer element and wait for appropriate inductive signals that potentiate its action on the *PEE* via chromatin looping. Another interesting finding in our study is the discrepancy of DNA binding affinity of the GATA2

mutants, particularly related to the *PEE* region. Firstly, L359V is a gain-of-function mutation in CML-BC and has been shown to enhance DNA binding affinity to target DNA. The reduced affinity to *Prox1* -353 probe by L359V fits with our previous HL-60 microarray data (see **Chapter 3**) in which the protein gains ability to actuate a further 1252 genes, but simultaneously lose the ability to regulate 446 genes. Secondly, T354M, R362Q and R398W bound to the *Prox1* -1352 and -353 probes at level similar to or slightly less than WT. Patients harbouring these mutations show no sign of lymphoedma, whereas patients that have been seen with lymphoedema harbour complete LOF mutation such as R361L and C373R. Together, these results suggest that GATA2 transactivation threshold (*e.g.* 50% activity in Emberger syndrome patients, see **Chapter 6** for more detail) is a critical determinant for the development of specific disease. We also predict that patients harbouring T355del might be predisposed to lymphoedema, as the mutant shares many common functional characteristics such as poor or no DNA binding and low transactivation activity with R361L and C373R. Numerous questions on how these mutants cause different phenotypes remain unanswered. Further investigations are warranted on this important question.

5.3.2 *Prox1* Enhancer is a Functioning *cis*-Regulatory Element

Computational data mining at the UCSC genome browser (position at chr1:214150377-214151377, GRCh37/hg19) revealed that the *Prox1* enhancer is highly conserved between human and mouse. The human *PROX1* -10kb homologous region also revealed that at least 7 cell lines, including GM12878 (lymphoblastoid cell lines), H1-hESC (H1 human embryonic stem cells), HSMM (human skeletal muscle myoblasts), HUVEC (human umbilical vein endothelial cells), K562 (human erythroleukemia cell line), NHEK (normal human epidermal keratinocytes), NHLF (normal human lung fibroblast) contain signatures for H3K27Ac. Distinctively from other inactive/poised enhancer elements, H3K27Ac is a signature for the active enhancer [30]. ChIP-Seq data from ENCODE show that the human *PROX1* -10kb homologous region is bound by multiple TFs including GATA2, c-JUN, c-FOS and TCF4 in non-LEC (*e.g.* HUVEC). Furthermore, we have recently demonstrated that at least 2 TFs (GATA2 and FOXC2) bind to this region in HUVEC and K562 cells (Jan Kazenwadel, personal communication). Together, these data implicate that the mouse *Prox1* -11.3kb fragment is likely a long-range regulatory element for *Prox1*. Intriguingly, WEMSA studies

showed that GATA2 mutant proteins bound differentially to the -11.3kb DNA probe. We speculate that GATA2 differential DNA binding affinity might plausibly be driven by individual differences of the mutants, ultimately, resulting in phenotypes attributed by them. This might occur via differential interactions with GATA2 binding partners as the *PEE* and -11.3kb enhancer regions are potentially activated by multiple TFs. Since multiple TFs were found to bind to human *PROX1* enhancer, we investigated GATA2's role as a co-activator for *Prox1* expression. Our data suggest GATA2 has an additive or synergistic regulatory effect with other TFs. Interestingly, all TFs except for PROX1 activated -11.3kb enhancer. We speculate that the inhibitory effect of PROX1 is triggered by a negative feedback action to regulate its own expression. This speculation is consistent with the finding that adenovirus mediated overexpression of PROX1 resulted in a significant downregulation of the endogenous PROX1 [31]. We may have overlooked or underestimated the complexity of genome-wide protein-DNA interactions, including TF binding and histone modifications *in vivo*. Under optimal conditions, the majority of enhancer/promoter regions are active when tested in reporter assays. To accurately survey protein-DNA interactions, future ChIP assays with each TF looking specifically at this region (either as a co-activator or repressor) with appropriate target cells (*e.g.* LEC) should be performed.

5.4 Conclusion

In conclusion, our data conjointly unveil an important regulatory role for GATA2 in the lymphatic compartment. Our study delineates two *Gata2* enhancer/promoter elements that can direct reporter gene expression *in vitro*. We also expand the spectrum of GATA2 mutations and their associated DNA binding affinity and transactivation on the *Prox1* promoter/enhancer. The field on how *GATA2* causes lymphoedema is still in its infancy. However, the data presented here improve our understanding and knowledge by which GATA2 mutation may contribute to lymphoedema in human.

5.5 Recommendations for Future Work

In this chapter, we reported the discovery of two putative GATA2 binding sites in the regulatory region of the *Prox1* locus. In addition, our WEMSA and luciferase assays showed that GATA2 reacts to both regulatory regions. Still, we do not know how GATA2 modulates *Prox1* expression in LEC. ChIP assays for GATA2 and other critical lymphatic TFs such as FOXC2 and SOX18 in LEC are currently being conducted. Another important question is: **What is the role of *GATA2* in the lymphatic system that leads to development of lymphoedema?** We hypothesize that complete LOF of GATA2 is associated with Emberger syndrome [8]. To specifically answer the question mentioned above, our collaborator, Dr Natasha Harvey has brought in *Gata2*$^{flox/flox}$ conditional KO mice from Finland [32]. Conditional activation of *Cre*-recombinase in *Gata2* $^{flox/-}$ mice can results in genomic deletion of one copy of *Gata2*, a genotype which resembles to *GATA2* mutations in human Emberger syndrome. Thus, this approach offers an ideal platform to understand the lymphatic vascular development and valve formation.

5.6 References

1. Schulte-Merker, S., A. Sabine, and T.V. Petrova, *Lymphatic vascular morphogenesis in development, physiology, and disease.* The Journal of cell biology, 2011. **193**(4): p. 607-18.
2. Karkkainen, M.J., et al., *Missense mutations interfere with VEGFR-3 signalling in primary lymphoedema.* Nature genetics, 2000. **25**(2): p. 153-9.
3. Fang, J., et al., *Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome.* Am J Hum Genet, 2000. **67**(6): p. 1382-8.
4. Irrthum, A., et al., *Mutations in the transcription factor gene SOX18 underlie recessive and dominant forms of hypotrichosis-lymphedema-telangiectasia.* Am J Hum Genet, 2003. **72**(6): p. 1470-8.
5. Slatter, D.H., *Textbook of Small Animal Surgery*. 2003: Saunders. p.1074.
6. Hennekam, R.C., et al., *Autosomal recessive intestinal lymphangiectasia and lymphedema, with facial anomalies and mental retardation.* Am J Med Genet, 1989. **34**(4): p. 593-600.
7. Alders, M., et al., *Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans.* Nature genetics, 2009. **41**(12): p. 1272-4.

8. Kazenwadel, J., et al., *Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature.* Blood, 2012. **119**(5): p. 1283-91.
9. Ostergaard, P., et al., *Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome).* Nature genetics, 2011. **43**(10): p. 929-31.
10. Ishida, H., et al., *GATA-2 anomaly and clinical phenotype of a sporadic case of lymphedema, dendritic cell, monocyte, B- and NK-cell (DCML) deficiency, and myelodysplasia.* Eur J Pediatr, 2012. **171**(8): p. 1273-6.
11. Lee, M.E., et al., *Cloning of the GATA-binding protein that regulates endothelin-1 gene expression in endothelial cells.* J Biol Chem, 1991. **266**(24): p. 16188-92.
12. Deleuze, V., et al., *TAL-1/SCL and its partners E47 and LMO2 up-regulate VE-cadherin expression in endothelial cells.* Mol Cell Biol, 2007. **27**(7): p. 2687-97.
13. Mollica, L.R., et al., *Role of a 5'-enhancer in the transcriptional regulation of the human endothelial cell protein C receptor gene.* Blood, 2006. **108**(4): p. 1251-9.
14. Kappel, A., et al., *Role of SCL/Tal-1, GATA, and ets transcription factor binding sites for the regulation of flk-1 expression during murine vascular development.* Blood, 2000. **96**(9): p. 3078-85.
15. Gumina, R.J., et al., *Characterization of the human platelet/endothelial cell adhesion molecule-1 promoter: identification of a GATA-2 binding element required for optimal transcriptional activity.* Blood, 1997. **89**(4): p. 1260-9.
16. Cowan, P.J., et al., *The human ICAM-2 promoter is endothelial cell-specific in vitro and in vivo and contains critical Sp1 and GATA binding sites.* J Biol Chem, 1998. **273**(19): p. 11737-44.
17. Chan, W.Y., et al., *The paralogous hematopoietic regulators Lyl1 and Scl are coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype.* Blood, 2007. **109**(5): p. 1908-16.
18. Khandekar, M., et al., *A Gata2 intronic enhancer confers its pan-endothelia-specific regulation.* Development, 2007. **134**(9): p. 1703-12.
19. Lim, K.C., et al., *Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning.* J Clin Invest, 2012. **122**(10): p. 3705-17.
20. Wilson, N.K., et al., *Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.* Cell Stem Cell, 2010. **7**(4): p. 532-44.
21. Francois, M., et al., *Sox18 induces development of the lymphatic vasculature in mice.* Nature, 2008. **456**(7222): p. 643-7.
22. Consortium, E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.
23. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
24. Pennisi, D., et al., *Mice null for sox18 are viable and display a mild coat defect.* Mol Cell Biol, 2000. **20**(24): p. 9331-6.
25. Oliver, G., *Lymphatic vasculature development.* Nat Rev Immunol, 2004. **4**(1): p. 35-45.
26. Bossard, P. and K.S. Zaret, *GATA transcription factors as potentiators of gut endoderm differentiation.* Development, 1998. **125**(24): p. 4909-17.

27. Gualdi, R., et al., *Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control.* Genes Dev, 1996. **10**(13): p. 1670-82.
28. Cirillo, L.A. and K.S. Zaret, *An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA.* Mol Cell, 1999. **4**(6): p. 961-9.
29. Cirillo, L.A., et al., *Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4.* Mol Cell, 2002. **9**(2): p. 279-89.
30. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
31. Yoo, J., et al., *Opposing regulation of PROX1 by interleukin-3 receptor and NOTCH directs differential host cell fate reprogramming by Kaposi sarcoma herpes virus.* PLoS Pathog, 2012. **8**(6): p. e1002770.
32. Haugas, M., et al., *Gata2 is required for the development of inner ear semicircular ducts and the surrounding perilymphatic space.* Dev Dyn, 2010. **239**(9): p. 2452-69.

Chapter 6: Final Discussion and Conclusion

6.1 GATA2 is a Predisposition Gene for MDS/AML, Infectious Diseases and Emberger Syndrome

Genetic factors have been long recognized as one of the major determinants of cancer. In fact, heritable cancers account for approximately 5-10% of all cancer cases [1]. In this thesis, we describe the discovery of the first heritable *GATA2* mutation T354M in 3 large families that are susceptible to MDS/AML and T355del mutation in 1 small family that predisposes to MDS [2, 3]. Systematic analyses on the 3 large pedigrees revealed that T354M is multi-generational with early onset of MDS/AML and aggressive courses. On the other hand, T355del was only identified in 2 individuals of a single small family. Both father and son presented MDS and have undergone BM transplantation. Since then no follow-up has been reported. Nonetheless, this finding indicates that some *GATA2* germline mutations may not be embryonic lethal and are compatible with life. However, patients harbouring T354M mutations have an increased risk of various HM, particularly MDS/AML, as well as a compromised immune system due to various different cytopaenias.

Summarization of published literature and in-house unpublished data on germline *GATA2* mutations and their clinical outcome shows that as well as the T354M and T355del mutations, there are 8 other reported germline mutations, namely R361L, C373R, R396Q, R396W and R398W (missense mutations in ZF2), R78Pfs*107 and L105Pfs*15 (premature termination mutations) [4] and a large deletion mutation, M1del290 [5, 6]. All mutations predispose affected individuals to MDS/AML, infectious diseases and cytopenias. R361L and C373R were previously reported in single patients with primary lymphoedema, MDS and HPV infection [4]. Our study indicates that both mutants are highly likely to be almost complete LOF with respect to DNA binding, which is consistent with the C373R mutation disrupting an essential cysteine in the C4 zinc ion coordinating ZF structure and R361 being a critical residue for WGATAR DNA sequence recognition. Thus, we conclude that this would generate

complete haploinsufficiency for GATA2 as seen for complete GATA2 gene deletion. We propose that such haploinsufficiency prediposes to lymphoedema. The other germline mutations are partial LOF (at least in some of our assays) that possibly exert their effects by altering the expression of critical GATA2 target genes or via aberrant protein-protein interactions. Moreover, recent studies have established that germline deletion or mutation of *cis*-acting auto regulatory elements located in 5'UTR, as well as the "GATA Switch" in intron 4 of *GATA2* causes a significant reduction in its own expression and predisposes to aplastic anaemia and immunodeficiency syndromes [7, 8]. Taken together, there is a role for reduced GATA2 function in leukaemic transformation whether by reduced expression of the normal protein or mutation of the protein.

To date, there have been a total of 52 T354M mutation cases reported (48 familial cases in 6 families, 1 *de novo*, 3 unknown). Reported patients with T354M mutation present diverse phenotypic spectrum include MDS (REAB-1, REAB-2, RA), MDS/AML, AML (predominantly M2, also M7), as well as aplastic anaemia, MonoMAC syndrome and DCML deficiency [3, 5, 9-13]. Interestingly, T354M (52 cases) shows no clinical signs of lymphoedema. This raises the question of what is the correlation between GATA2 expression/activity level and phenotypic presentation. Perhaps studies of GATA3 could give us a glimpse into the enigma of T354M. Human GATA3 haploinsufficiency is known to lead to HDR (hypoparathyroidism, deafness and renal dysplasia) syndrome. Studies showed that defects in kidney development in mice can be restored with at least 70% of diploid GATA3 activity [14]. Another study demonstrated that GATA3 T272I mutant which has approximately 30% of the WT DNA binding function, is associated with parathyroid, otic, and renal abnormalities [15]. We speculate that the similar phenomenon also occurs for GATA2. It is possible that for T354M, the 75% of GATA2 activity is just enough to protect T354M patients from lymphoedema, but is still permissive to predispose to MDS/AML and/or immunodeficiency. Furthermore, our WEMSA assays showed that those mutants (*e.g.* L359V, R362Q, R398W) that bind DNA as well as or better than T354M also do not exhibit clinical signs of lymphoedema, but do allow MDS/AML and immunodeficiency. However, the complete or partial deletion mutants or framshift mutants or those which bind DNA poorly (*e.g.* R361L and C373R; where GATA2 activity is predicted to be only ≈50%) are all seen to

predispose to lymphoedema. Together, these data suggest that GATA2 activity levels could provide a critical threshold for specific disease manifestation. In addition, it is highly likely that environment factors or pressures (*e.g.* endemic Mycobaterial pathogens, high rate of HPV transmission, poor hygiene or lymphatic stresses) place stress on an inherently weakened or prone haematopoietic or lymphatic system. Hence, ultimate clinical phenotypic variation is likely attributable to both genetic changes plus the environment effects. In summary, T354M significantly affects the balance of sequestration of cofactors, as well as the expression of GATA2 target genes. Together, these findings point towards T354M affecting and altering haematopoietic homeostasis, resulting in a dysregulated network of haematopoietic TFs, their target genes, and cellular processes, including renewal, expansion and differentiation culminating in MDS/AML or immunodeficiency.

The rest of the 3 germline mutants, R396Q (12 patients - 4 family, 2 *de novo* patient) [5, 11, 16-18], R396W (2 *de novo* patients) [5, 6, 19] and R398W (9 patients - 6 families) [5, 6, 10, 20] have been reported to have mainly in MonoMAC/DCML deficiency syndrome (14 cases), MDS and/or AML (8 cases) and, lymphoedema and immunodeficiency (1 case) (see **Figure 6.1**). Germline mutations can either emerge *de novo* or originate from a common founder mutation (familial). The multiple recurrent germline mutations at specific positions in GATA2 implicate that ZF2 probably is a hotspot for repeated *de novo* mutation. There have been reports of protein-driven positive selection for germ cells. Goriely *et al.* demonstrated that some GOF mutations (*e.g.* S252W and S252L) in fibroblast growth factor receptor 2 (FGFR2) are positively selected in spermatogonial cells because they encode proteins with GOF properties favouring clonal expansion [21].

However, because of the rarity of GATA2 mutant families or *de novo* cases, we do not envisage gonadal selection for GATA2 mutants. Rather, the data from reported literature is consistent with these mutations arising sporadically either in germ cells or the one cell embryo, and those that are not embryonic lethal and compatible with life and cause a major predisposition are ultimately seen. Hence, this could lead to the occurrence of *GATA2 de novo* mutations either in an individual case or as a founder mutation in affected families as seen for T354M. Interestingly, our analyses of the literature together with our data implicate that there

is a strong genotype-phenotype correlation for GATA2 mutations. For instance, patients harbouring germline GATA2 mutations exhibited an array of common phenotypes (immunodeficiency and/or MDS/AML and/or lymphoedema). In contrast, none of the patients with somatic GATA2 mutations presented with immunodeficiency, MDS/AML or lymphoedema. They often presented with a specific phenotype, either the *CEBPA* (mono- and biallelic) AML or AML-M5. In summary, positive germline selection [22] and recurrent specific residue mutations with an autosomal dominant inheritance fashion [23] are highly suggesting that T354M, R396Q, R396W and R398W are GOF mutations.

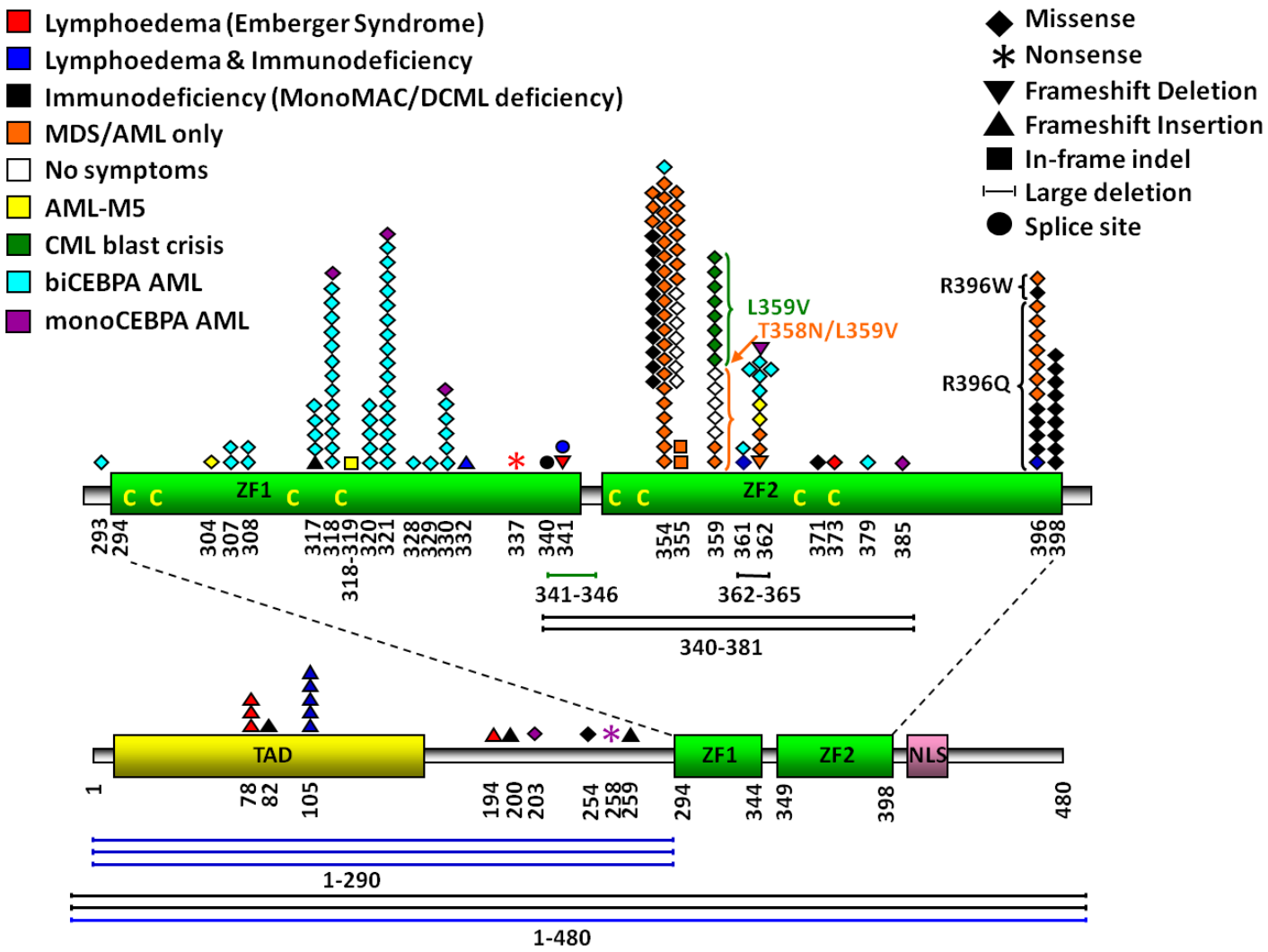


Figure 6.1: Distribution of GATA2 Mutations Identified in Various Haematological Malignancies, MonoMAC/DCML Deficiency and Emberger Syndrome. Each symbol represents an independent mutation. Abbreviations: TAD, transactivation domain; C, Zinc coordinating cysteine; ZF1, zinc finger 1; ZF2, zinc finger 2; NLS, nuclear localization signal.

6.2 GATA2 in Sporadic AML and CML

We have reviewed and compiled all published *GATA2* somatic mutations and their phenotypes. Careful analyses of the data revealed that *GATA2* mutations are rare in sporadic AML [3, 24, 25]. Nonetheless, we cannot rule out that the screening strategy and low tumour load may have contributed to missing or masking possible GATA2 mutations. Interestingly, several reports have shown that somatic *GATA2* mutations do occur, but in specific AML subtypes. Somatic *GATA2* mutations are detected in a small number of AML-M5 (3.6%; 4/112) [26] and in cytogenetically normal (CN) AML with biallelic *CEBPA* mutations (24.7%; 48/186) [27-29]. Very recently, Green *et al.* described the co-occurrence of GATA2 mutations in CN AML patients with monoallelic *CEBPA* mutations (16.3%; 7/43) [29] as opposed to Greif *et al.* [27] and Fasan *et al.* [28] (0%; 0/60) who did not find any GATA2 mutations. Another report described 5 GATA2 mutant cases of CN paediatric AML (unknown germline status) with other co-occurrent mutations, 3 of which had different amino acid substitution (R362Q/P/G) at the same position [30]. These results suggest that GATA2 somatic mutations may occur in specific AML subtypes (as opposed to T354M) and co-occur with various other recurrent mutations such as biallelic *CEBPA, NRAS, WT1, RUNX1* and *NPM1* [28, 30].

GATA2 mutations have been described in *de novo* AML patients with biallelic *CEBPA* mutations [27-29]. Genotypic analysis revealed that the majority of these mutations cluster within ZF1, particularly at residues 317-321, suggesting that this region is likely a mutational hotspot for GATA2 in the biallelic *CEBPA* mutant AML sub-group (see **Figure 6.1**). Of note, a *CEBPA* mutation is not a mandatory condition to attain a *GATA2* mutation, as studies have shown that *GATA2* mutations do occur in *CEBPA* WT AML, although at a lower frequency (only 6 cases reported) [28, 30]. Intriguingly, *GATA2* mutations in biallelic *CEBPA* AML patients disappeared as expected during remission, but were not seen in relapse samples (3 out of 4), indicating that *GATA2* mutations are secondary events in this sub-group [28]. This data was soon reinforced by the first report of familial *CEBPA* AML, in which a second allelic *CEBPA* and GATA2 (2 monoallelics and 1 biallelic) mutations were both acquired by 3 family members [29]. In particular, all 3 family members acquired missense mutations in the ZF1 domain of GATA2, a pattern similar to that of seen in the somatic biallelic *CEBPA* AML [29].

Together, these data highly suggest that this group of GATA2 mutations represents a unique entity that is segregated from other germline mutations such as T354M. Another unique feature for biallelic *CEBPA* mutations is the co-occurrence of biallelic *GATA2* mutations, probably on different alleles (of 5 reported cases, 2 have been confirmed biallelic *GATA2*) [27, 29]. Notably, one mutation is located in the ZF1 hotspot zone (317-321), while the other mutation can be elsewhere in the protein (see **Figure 6.2** for #). The reason for this phenomenon (biallelic *GATA2* mutations), however, is not known. We also surmise that GATA2 mutations do occur randomly in AML with *CEBPA* mutations. Indeed, the relatively high incidence of GATA2 mutations co-occurring with *CEBPA* mutations (mono- or biallelic) suggests that there is a strong positive selection process for these GATA2 mutations, particularly for the mutations in ZF1. These functionally abnormal GATA2 are predicted to confer an extra neoplastic advantage for the *CEBPA* mutated cells. Greif *et al*. demonstrated that GATA2 ZF1 mutant bound WT CEBPA less well than WT GATA2. Hence, it is tempting to speculate that in the unique biallelic CEBPA mutant situation (*i.e.* one N-terminal frameshift mutation and one C-terminal in-frame indel mutation), WT GATA2 binds and sequesters to the mutant CEBPA proteins which, upon mutation of the GATA2 ZF1, are then released to more, strongly drive the oncogenic process. It also follows then that a second GATA2 mutation on the other allele could occur to release even more mutant CEBPA protein leading to the biallelic CEBPA and biallelic GATA2 situation.

The rarity of biallelic *GATA2* mutations in sporadic AML, implicates that *GATA2* does not fit into the classical LOF tumour suppressor gene model. One likely explanation is that AML stem cells must retain some residual GATA2 function that is essential for leukaemic cell survival or selection and hence heterozygous *GATA2* mutation (haploinsufficiency). And perhaps the most compelling of our current analysis is that the majority of the somatic mutations (*e.g.* AML-M5 and CN-AML with mono- or biallelic *CEBPA* mutations) cluster in ZF1 (except for R361, R362, L379 and P385) and recurrent germline mutations occur mostly in ZF2 rather than the classical randomly spread frameshift, nonsense and missense mutations characteristic of most tumour suppressor mutations. Intriguingly, GATA2 amino acid residues 377-475 have been shown to directly contact CEBPA [31] and 2 non-ZF1 mutations are in this region (L379Q and P385L) further supporting for GATA2 and CEBPA a collaborative role

mutations in certain subtypes of AML. We do not observe any overlap in missense mutations for confirmed somatic mutations and germline mutations (except for R361L) (see **Figure 6.2**). As aforementioned, somatic GATA2 mutations are a secondary event that occurs predominantly in specific AML sub-types. Although they may confer an additional advantage that allows them to be selected in "already" mutated cells, these ZF1 mutations are virtually not seen in the familial or *de novo* AML. The unique segregation pattern observed in GATA2 mutations is highly suggestive that these mutations are not compatible with life (as opposed to the germline mutations) and possibly play a cooperating role in the initiation of leukaemogenesis. Hence, they represent a clinically and biologically unique disease entity.

In contrast, *L359V* is a recurrent somatic mutation that has not yet been reported in any HM other than CML-BC [24, 25]. CML-BC, clinically resembles AML, although it is commonly characterized by specific BCR-ABL translocation (Philadelphia chromosome; Ph) [32]. Oncogenic activation in CML, to progress from the chronic phase to blast phase, requires additional somatic genetic event(s). The Ph, itself can drive genetic instability and may contribute to GATA2 mutation. The question is, why L359V is recurrently mutated in CML-BC and not other GATA2 mutations? In this situation, it seems that selection is operating on L359V. Based on the functional assays and microarray data, L359V is, in some ways, the "odd-one-out", being both a partial LOF and a significant GOF at the DNA binding and transactivation levels. We propose that it interferes with differentiation of the Ph^+ clones, hence, driving the CML chronic phase (with rapidly dividing but differentiating cells) to accelerated phase or blast phase (rapidly dividing, but blocked in differentiation).

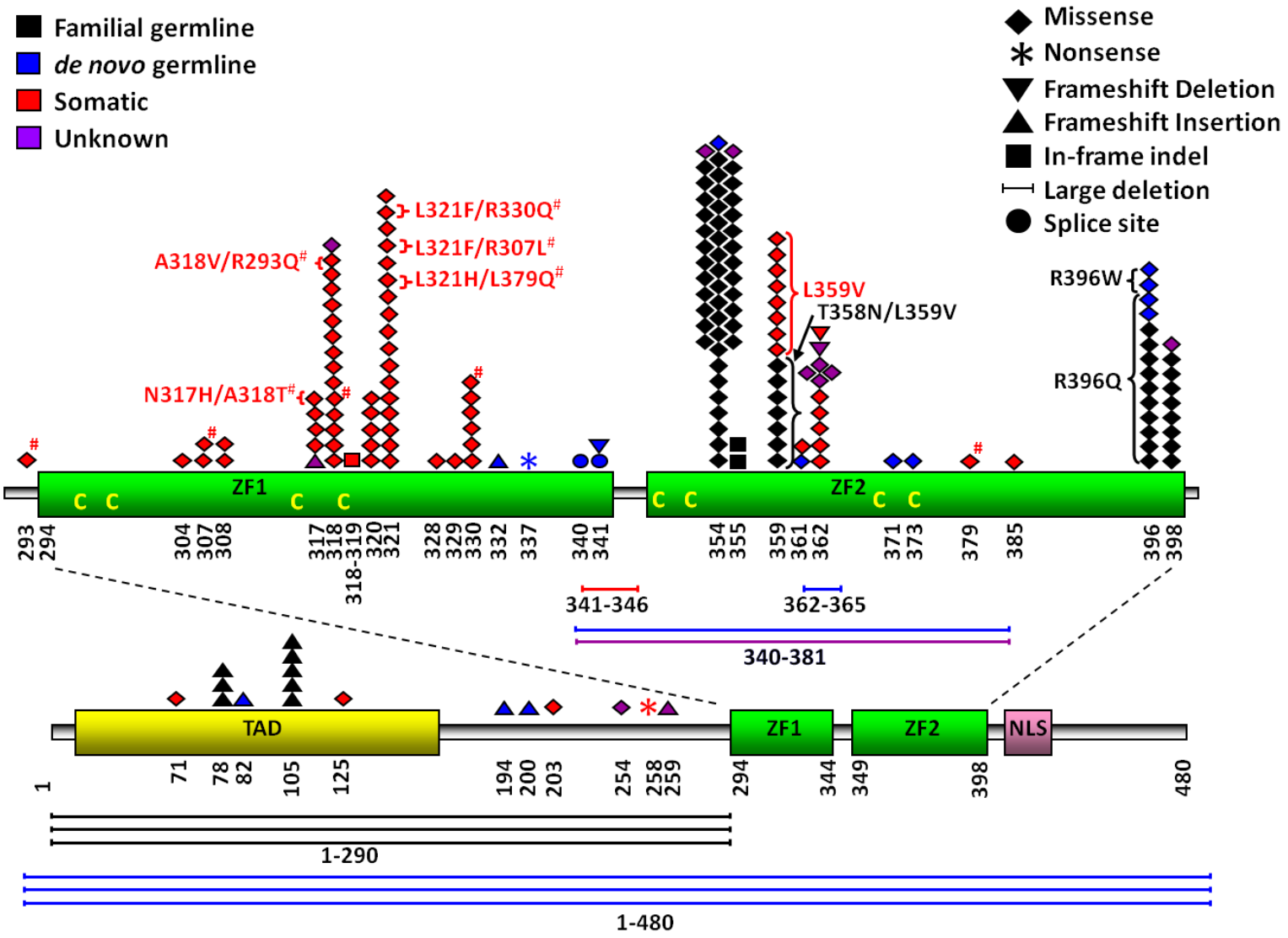


Figure 6.2: GATA2 Mutant Proteins According to Germline or Somatic Mutation. # indicates the mutated residues in patients who also harbour a second GATA2 mutation.

6.3 Significance and Impact of the GATA2 Discovery

We first report the discovery of *GATA2* mutations that predisposes to MDS/AML in one Australian family and 3 American families. Although significant progress has been made in delineating GATA2 mutations related diseases, these patients generally do not respond very well to conventional chemotherapy and died from the diseases or other complications [9, 17]. Realizing the "nasty" outcome of GATA2 mutations and the ineffectiveness of chemotherapy, Cuellar-Rodriguez *et al.* performed HSC transplantations in 6 patients with MDS and MonoMAC/DCML deficiency syndrome [33]. Remarkably, 5 out of 6 patients were apparently cured and demonstrated reversal of the haematologic, immunologic, and clinical manifestations, implicating that BM transplantation is a viable and effective option for these patients. In Adelaide, genetic testing for GATA2 coding mutation is available through diagnostic services at SA Pathology/IMVS in the Department of Genetic Pathology, and genetic counselling is available by the clinical geneticists. Individuals with positive GATA2 mutation disease can be monitored for tumour load, or can plan for BM or HSC transplantation with a HLA matched mutation negative sibling or relative donor in the future. To date, at least 4 unaffected individuals from the T354M Australian pedegree have made appointments and been tested for GATA mutations with all being negative. Moreover, a local patient with MDS and multiple warts on the palms and knees was symptomatically predicted by a local haematologist to carry a GATA2 mutation, and upon verification of a splice site mutation, the patient was promptly BM transplanted. Remarkably, the warts that presented on the patient have been cured after transplantation, as well as the MDS, demonstrating that the phenotype is reversible and curable (personal communication with Dr. David Yeung and Dr. Peter Bardy, Haematologists). The discovery of GATA2 mutations and its clinical outcome is an example of how fundamental research can be translated and provide benefits for clinical and public health applications in the form of improved risk assessment, early detection, more precise diagnosis and prognosis prediction, disease prevention and targeted therapeutics [34].

6.4 Conclusion

The information gleaned from the studies presented in this thesis has provided new insights into the genetic basis and function for the multi patho-phenotypic divergence that is associated with recurrent point mutations in *GATA2*. The study of *GATA2* mutation will open up opportunities to explore the mechanism on disease initiation and progression. However, further studies and efforts are still required to better understand the mechanism contributed by each of the *GATA2* mutations. We are begun to dissect out the mechanisms leading to these states. Through the course of this thesis, the discovery of GATA2 in myeloid and lymphatic diseases has opened up a whole new field of discovery. Different mutations drive different phenotypes and we are beginning to dissect out the mechanisms leading to these states. By deciphering the pathogenicity of GATA2 mutants at the molecular level, we may unveil known and novel biochemical pathways that lead to cancer development, opening the way for alternative therapies for GATA2-related malignancy.

6.5 References

1. Macdonald, F., C. Ford, and A. Casson, *Molecular Biology of Cancer*. 2004: Taylor & Francis. p.77.
2. Scott, H.S., et al. *GATA2 is a new predisposition gene for familial myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML) [Late-breaking Abstracts]*. in *American Society of Hematology Annual Meeting*. 2010. Orlando Florida.
3. Hahn, C.N., et al., *Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia.* Nature genetics, 2011. **43**(10): p. 1012-7.
4. Ostergaard, P., et al., *Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome).* Nature genetics, 2011. **43**(10): p. 929-31.
5. Hsu, A.P., et al., *Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome.* Blood, 2011. **118**(10): p. 2653-5.
6. Vinh, D.C., et al., *Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia.* Blood, 2010. **115**(8): p. 1519-29.

7. Townsley, D.M., et al., *Regulatory Mutations in GATA2 Associated with Aplastic Anemia. ASH Abstract. Nov 2012; 120: 3488.* , in *American Society of Hematology Annual Meeting*2012: Atlanta, GA.
8. Johnson, K.D., et al., *Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity.* J Clin Invest, 2012. **122**(10): p. 3692-704.
9. Bodor, C., et al., *Germ-line GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival.* Haematologica, 2012. **97**(6): p. 890-4.
10. Dickinson, R.E., et al., *Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency.* Blood, 2011. **118**(10): p. 2656-8.
11. Holme, H., et al., *Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia.* Br J Haematol, 2012. **158**(2): p. 242-8.
12. Kazenwadel, J., et al., *Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature.* Blood, 2012. **119**(5): p. 1283-91.
13. Mace, E.M., et al., *Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56bright subset.* Blood, 2013.
14. Hasegawa, S.L., et al., *Dosage-dependent rescue of definitive nephrogenesis by a distant Gata3 enhancer.* Dev Biol, 2007. **301**(2): p. 568-77.
15. Gaynor, K.U., et al., *A missense GATA3 mutation, Thr272Ile, causes the hypoparathyroidism, deafness, and renal dysplasia syndrome.* J Clin Endocrinol Metab, 2009. **94**(10): p. 3897-904.
16. Ishida, H., et al., *GATA-2 anomaly and clinical phenotype of a sporadic case of lymphedema, dendritic cell, monocyte, B- and NK-cell (DCML) deficiency, and myelodysplasia.* Eur J Pediatr, 2012. **171**(8): p. 1273-6.
17. Pasquet, M., et al., *High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia.* Blood, 2013. **121**(5): p. 822-9.
18. Alter, B.P., et al., *Clinically Silent Carriers in Families with Myelodysplastic Syndrome Due to GATA2 Mutations ASH Abstract. Nov 2012; 120: 1264.* , in *American Society of Hematology Annual Meeting*2012: Atlanta, GA.
19. Cellot, S. and P. Teira, *Unpublished GATA2 mutation*, 2012: Sainte Justine Hospital, Montreal, Quebec, Canada.
20. Bigley, V., et al., *The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency.* The Journal of experimental medicine, 2011. **208**(2): p. 227-34.
21. Goriely, A., et al., *Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia.* Proc Natl Acad Sci U S A, 2005. **102**(17): p. 6051-6.
22. Arnheim, N. and P. Calabrese, *Understanding what determines the frequency and pattern of human germline mutations.* Nat Rev Genet, 2009. **10**(7): p. 478-88.
23. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
24. Zhang, S.J., et al., *Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia.* Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2076-81.

25. Zhang, S.J., J.Y. Shi, and J.Y. Li, *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.* Leuk Res, 2009. **33**(8): p. 1141-3.
26. Yan, X.J., et al., *Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia.* Nature genetics, 2011. **43**(4): p. 309-15.
27. Greif, P.A., et al., *GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia.* Blood, 2012. **120**(2): p. 395-403.
28. Fasan, A., et al., *GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis.* Leukemia, 2013. **27**(2): p. 482-5.
29. Green, C.L., et al., *GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations.* Br J Haematol, 2013.
30. Luesink, M., et al., *High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia.* Blood, 2012. **120**(10): p. 2064-75.
31. Tong, Q., et al., *Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation.* Mol Cell Biol, 2005. **25**(2): p. 706-15.
32. Rowley, J.D., *Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining.* Nature, 1973. **243**(5405): p. 290-3.
33. Cuellar-Rodriguez, J., et al., *Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency.* Blood, 2011. **118**(13): p. 3715-20.
34. Schully, S.D., et al., *Translational research in cancer genetics: the road less traveled.* Public Health Genomics, 2011. **14**(1): p. 1-8.

APPENDIX A: Additional Information for Chapter 3

A-1 STATEMENT OF AUTHORSHIP

Title of Paper	**HERITABLE GATA2 MUTATIONS ASSOCIATED WITH FAMILIAL MYELODYSPLASTIC SYNDROME AND ACUTE MYELOID LEUKEMIA**
Publication Status	• Published ○ Accepted for Publication ○ Submitted for Publication ○ Publication Style
Publication Details	Christopher N Hahn, **Chan-Eng Chong**, Catherine L Carmichael, Ella J Wilkins, Peter J Brautigan, Xiao-Chun Li, Milena Babic, Ming Lin, Amandine Carmagnac, Young K Lee, Chung H Kok, Lucia Gagliardi, Kathryn L Friend, Paul G Ekert, Carolyn M Butcher, Anna L Brown, Ian D Lewis, L Bik To, Andrew E Timms, Jan Storek, Sarah Moore, Meryl Altree, Robert Escher, Peter G Bardy, Graeme K Suthers, Richard J D'Andrea, Marshall S Horwitz & Hamish S Scott.

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's (***Chan Eng Chong***) thesis.

Name of Co-Author	Christopher N. Hahn		
Contribution to the Paper	Transactivation assays, supervised development of the works and acted as first author		
Overall percentage (%)	20		
Signature		Date	30/4/2013

Name of Co-Author	Chan-Eng Chong		
Contribution to the Paper	Performed all tissue cultures (HEK293, HEK293T and HL-60), protein and RNA extractions, transduction of GATA2 into HL-60, proliferation assays, differentiation assays, immunofluorescence staining, construct generation, mutagenesis, morphology studies (cytospin), WEMSA, EMSA and data analysis on all samples, interpreted data and acted as co-author (second author).		
Overall percentage (%)	20		
Signature		Date	30/4/13

Name of Co-Author	Catherine L Carmichael		
Contribution to the Paper	Choose 50 genes for sequencing and analyzed gene sequence		
Overall percentage (%)	2		
Signature		Date	9/05/2013

Name of Co-Author	Ella J Wilkins		
Contribution to the Paper	Analyzed 50 gene sequences		
Overall percentage (%)	2		
Signature		Date	15/5/13.

Name of Co-Author	Peter J Brautigan		
Contribution to the Paper	Apoptosis assays and luciferase assays		
Overall percentage (%)	2		
Signature		Date	30/04/2013

Name of Co-Author	Xiao-Chun Li		
Contribution to the Paper	Generated original T354M mutation		
Overall percentage (%)	2		
Signature		Date	30/4/2013

Name of Co-Author	Milena Babic		
Contribution to the Paper	Sequencing and HRM		
Overall percentage (%)	2		
Signature		Date	30/04/13

Name of Co-Author	Ming Lin		
Contribution to the Paper	Generated lentiviral constructs and transduction		
Overall percentage (%)	2		
Signature		Date	30/04/13

Name of Co-Author	Amandine Carmagnac		
Contribution to the Paper	Analysed 50 gene sequences		
Overall percentage (%)	2		
Signature		Date	6/05/13

Name of Co-Author	Young K Lee		
Contribution to the Paper	WEMSA and western blots		
Overall percentage (%)	2		
Signature		Date	30/4/13.

Name of Co-Author	Chung H Kok		
Contribution to the Paper	Analysed microarray analysis		
Overall percentage (%)	2		
Signature		Date	30/4/13

Name of Co-Author	Lucia Gagliardi		
Contribution to the Paper	Analysed microarray analysis		
Overall percentage (%)	2		
Signature		Date	6/5/2013

Name of Co-Author	Kathryn L Friend		
Contribution to the Paper	Haplotype analysis		
Overall percentage (%)	2		
Signature		Date	1/5/13

Name of Co-Author	Paul G Ekert		
Contribution to the Paper	Advise on and provision of lentiviral constructs		
Overall percentage (%)	2		
Signature		Date	6/5/2013

Name of Co-Author	Carolyn M Butcher		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	30/4/13

Name of Co-Author	Anna L Brown		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	1/5/13

Name of Co-Author	Ian D Lewis		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	30/4/13

Name of Co-Author	L Bik To		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	24.2013

Name of Co-Author	Andrew E Timms		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	3/5/2013

Name of Co-Author	Jan Storek		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	3-May-2013

Name of Co-Author	Sarah Moore		
Contribution to the Paper	Provision of familial and sporadic MDS/AML		
Overall percentage (%)	2		
Signature		Date	30/4/13

Name of Co-Author	Meryl Altree		
Contribution to the Paper	Collection of familial samples and clinical histories		
Overall percentage (%)	2		
Signature		Date	3/5/13

Name of Co-Author	Robert Escher		
Contribution to the Paper	Selection of 50 AML genes		
Overall percentage (%)	2		
Signature		Date	3/5/13

Name of Co-Author	Peter G Bardy		
Contribution to the Paper	Access to familial samples and clinical data		
Overall percentage (%)	2		
Signature		Date	10.7.13

Name of Co-Author	Graeme K Suthers		
Contribution to the Paper	Access to familial samples and clinical data		
Overall percentage (%)	2		
Signature		Date	30/4/13

Name of Co-Author	Richard J D'Andrea		
Contribution to the Paper	Familial and sporadic MDS/AML and advice on functional assays		
Overall percentage (%)	4		
Signature		Date	30/4/13

Name of Co-Author	Marshall S Horwitz		
Contribution to the Paper	Familial and sporadic MDS/AML and advice on functional assays		
Overall percentage (%)	4		
Signature		Date	05/03/2013

Name of Co-Author	Hamish S Scott		
Contribution to the Paper	Familial and sporadic MDS/AML and advice on functional assays		
Overall percentage (%)	5		
Signature		Date	30/4/13

DECLARATION OF CONTRIBUTION

I also acknowledge the important contribution of Prof. Jacqueline Matthews and Dr Philippa Stokes (Protein structure, function and engineering laboratory at the University of Sydney) to the results presented in **Chapter 3** (Additional results). I assess my contribution to the results in this chapter to be as 10%.

A-2 SUPPLEMENTARY METHODS

Patients

Families (**Supplementary Table 2, Supplementary Note**) were recruited and sample use approved through institutional human ethics review board approved protocols from the Australian Familial Haematological Cancer Study (Royal Adelaide Hospital (RAH) #091203 and #100702, and Children, Youth and Women's Health Service #REC1542/12/12, Adelaide, SA Australia), The Queen Elizabeth Hospital and the University of Washington (Seattle, WA USA).

Sequence analysis of candidate genes

To identify germline and somatic mutations in patients with familial AML, a panel of 50 hematopoietic candidate genes, incorporating a total of 638 exons, was assembled (**Supplementary Table 1**). Primer design, PCR amplification, and dideoxy sequencing of genomic DNA purified from lymphoblastoid cells of probands from 7 MDS/AML pedigrees were performed by the Australian Genome Research Facility (AGRF). Sequences were aligned with NCBI RefSeq sequences using Mutation Surveyor (SoftGenetics) and variants compared to the UCSC and NCBI SNP databases for novelty. Sequence changes were confirmed by re-sequencing in both directions. Primer sequences are available upon request. Screening of control and sporadic AML populations was performed using high resolution melt (HRM) analysis (**Supplementary Table 8 and Supplementary Note**).

Cell culture

HEK293, 293T and Cos-7 cells were cultured in DMEM with 10% fetal bovine serum (FBS) (JRH Biosciences) and transient transfections were performed using Lipofectamine 2000 (Invitrogen). HL-60 promyelocytic cells were cultured in RPMI containing 10% FBS. All cultures contained 50 units/ml penicillin and 50 μg/ml streptomycin (Sigma).

Generation of mutant GATA2 plasmid and lentiviral expression constructs

An expression clone (pCMV6-XL6-GATA2) containing a 3.7 kb GATA2 cDNA insert was obtained from OriGene, and p.Thr354Met, p.Thr355del and p.Leu359Val mutants were

generated by site directed mutagenesis. The coding regions of wildtype (WT), p.Thr354Met and p.Leu359Val were cloned into a dual lentiviral vector system which was used to generate HL-60 cells expressing GATA2 WT or mutants upon addition of 4-hydroxytamoxifen (4HT) (see **Supplementary Note**).

GATA2-responsive promoter and enhancer studies

The GATA2-responsive promoter (*LYL1*) and enhancer (*RUNX*) were PCR amplified and cloned into pGL4.12[*luc2CP*] (*Sfi*I) and pGL3-Promoter (*Kpn*I/*Bgl*II) (Promega), respectively. The CSF1R (M-CSF-R) promoter was PCR amplified and cloned into pGL4.12[*luc2CP*] (*Sfi*I). See **Supplementary Table 9** for PCR primers used. The GATA2-responsive CD34 enhancer-luciferase construct (CD34x2/Luc) and one with the GATA binding sites mutated (mutant CD34x2/Luc) [2] were kindly provided by Tariq Enver, Weatherall Institute of Molecular Medicine, Oxford, U.K. HEK293 or Cos-7 cells were transfected at 90% confluence with Lipofectamine 2000. In all experiments, the molar equivalents of EV constructs were used to balance gene expressing constructs to avoid squelching artifacts. After 20 h, cells were harvested and luciferase activity determined with the Dual-Luciferase Reporter Assay System (Promega) using a GloMax®-Multi Detection System (Promega). All assays were performed a minimum of three times in triplicate. All results were analysed using Student's t-test, and reported as mean ± s.e.m. with significance, $p<0.05$ (asterisk).

Cell differentiation and proliferation assays

HL-60 cells were plated at 1.25 × 104 cells/ml and treated with or without 30 nm 4-hyrdoxy tamoxifen (4-HT) for 24 h and then with or without 2 μM all-trans retinoic acid (ATRA). Cell numbers were determined by manual counting and FACS analysis (Phycoerythrin anti-mouse CD11b and Phycoerythrin rat IgG2b isotype control) (eBioscience) was performed 6 days after addition of ATRA (Sigma). The cells were also stained with hematoxylin and eosin for assessing morphological changes.

Haplotyping

Haplotype mapping was performed by PCR amplification and sequencing of amplicons containing 50 single nucleotide polymorphisms (SNP) within and surrounding the position of the p.Thr354Met variant of the GATA2 gene (**Supplementary Table 10**). All amplicons were generated using AmpliTaq Gold (Applied Biosystems) according to the manufacturer's protocol using 2 mM MgCl2 and the following cycle strategy; 95°C, 10 min; 95°C, 30 s, 66°C – 58°C, 20 s (touchdown, 0.8°C/cycle for 10 cycles), 72°C, 45 s (total of 40 cycles); 72°C, 3 min.

Generation of mutant GATA2 plasmids and lentiviral expression constructs

An expression clone (pCMV6-XL6-GATA2) containing a 3.7 kb GATA2 cDNA insert was obtained from OriGene (Cat. No. SC125368). p.Thr354Met, p.Thr355del and p.Leu359Val mutants were generated by QuikChange mutagenesis (Stratagene) using the primers T354M-F and T354M-R, 355delT-F and 355delT-R, and L359V-F and L359V-R (**Supplementary Table 11**), respectively. For the generation of lentiviral expression constructs, the regulatable pF 5xUAS W SV40 Puro (5xUAS) [3] was used. GATA2 WT or mutants were PCR amplified from the above pCMV6 plasmid vectors using the primers (KOZAK-GATA2-F and either GATA2-FLAG-R or GATA2-R) (**Supplementary Table 12**) and *Pfu* Turbo (Stratagene), excised with *Xba*I and cloned into the unique *Xba*I site of 5xUAS.

Generation of regulatable GATA2 expressing HL-60 cell lines

A dual lentiviral vector system was used to generate HL-60 cells expressing GATA2 WT or mutants upon addition of 4-hydroxytamoxifen (4HT) (Sigma). Infectious third generation lentivirus was made by cotransfecting 293T cells with either 5xUAS-GATA2 (WT or mutants) or pF GEV16 Super PGKHygro (GEV16) [4] plasmid and the three packaging plasmids pHCMVwhvgagpolml, pHCMV-G and pHCMVwhvrevml [5] (mass ratio 50:5:2.5:1). Supernatants were harvested 24 h later and filtered (Nalgene 45 μm syringe filter) (Nalge Nunc Int.). HL-60 cells were firstly transduced with GEV16 lentiviral supernatant including 4 μg/ml polybrene and 2.5 μg/ml fungizone. After 48 h, HL-60GEV cells were selected in 1 mg/ml hygromycin (Roche). These cells were subsequently transduced with the GATA2 (WT,

p.Thr354Met, p.Thr355del and p.Leu359Val) or EV lentiviral supernatant and selected in 3 μM Puromycin (Sigma).

Immunofluorescence staining

HL-60 cells carrying stably transduced 4HT-regulatable GATA2 (WT, p.Thr354Met, p.Thr355del and p.Leu359Val) were treated with and without 100nM 4HT. After 24 h, the cells were fixed with 4% of paraformaldehyde for 10 min. The cells were permeabilized with 0.1% Triton/PBS, for 10 min and blocked with 2% BSA for 30 min. The cells were then stained with rabbit α-GATA2 antibody (Santa Cruz Biotechnology, Inc) (1:1000) for 1 h followed by Alexa 594-conjugated goat anti-rabbit secondary antibody (Molecular Probes) (2 μg/ml) for 20 min. The slides were mounted in Vectashield® mounting medium with DAPI (Vector Laboratories, Inc). Cells without primary antibody served as negative controls. All incubations were performed at room temperature.

Western blot analysis

HL-60 cells carrying stably transduced 4HT-regulatable GATA2 (WT, p.Thr354Met, p.Thr355del and p.Leu359Val) were treated with and without 100 nM 4HT. After 24 h, the cells were harvested in RIPA buffer (50 mM Tris-Cl pH 7.6, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate with protease inhibitor (cOmplete Mini EDTA free protease inhibitor tablets, Roche Diagnostics)). Samples were loaded onto the 10% acrylamide gels, electrophoresed and transferred onto Hybond-P PVDF membranes (Amersham). Membranes were probed with antibodies using standard techniques and visualised using ECL plus detection reagents (Amersham) on x-ray film (Amersham Hyperfilm™ MP).

Apoptosis Assays

HL-60 cells were stained for surface Annexin V and propidium iodide according to the manufacturer's protocol (#556547, Becton Dickinson).

Electromobility shift assay (EMSA) and EMSA-Western Blot

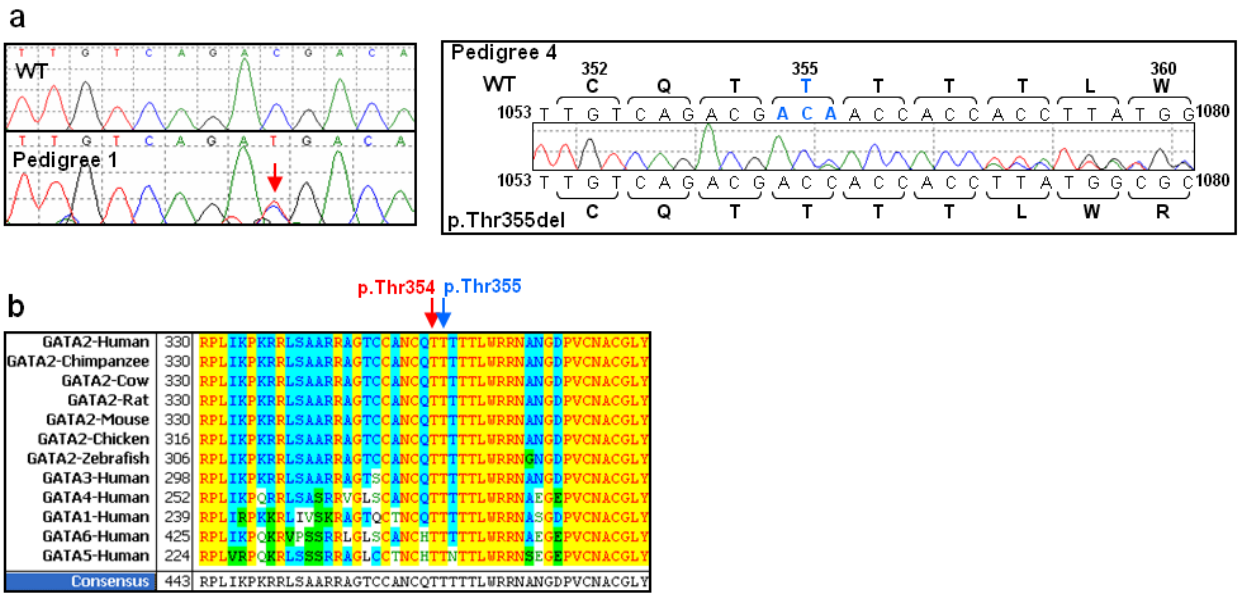
HEK293 cells were transfected with GATA2 WT or mutants using Lipofectamine™ 2000. After 24 h, nuclear extracts were prepared using a NE-PER® Nuclear and Cytoplasmic Extraction kit (Pierce) according to the manufacturer's protocol. Double stranded DNA oligonucleotides containing two GATA binding sites (Human TCRδ enhancer) or a single GATA binding site (GATA Consensus and Human GM-CSF-153 promoter) were synthesized (**Supplementary Table 13**). Each single stranded oligomer was labeled using a Biotin 3' End DNA Labeling kit (Pierce) and annealed according to manufacturer's protocol. Electrophoretic mobility shift assays were performed using a modified protocol from Kumar *et al* 2008 [6] and visualized using a Chemiluminescent Nucleic Acid Detection Module (Pierce) according to the manufacturer's protocol. Double stranded labeled probes (100 fmol) were incubated with 3 μg of nuclear extract for 20 min in 1x binding buffer containing 20 mM HEPES-KOH, pH 7.9, 100 mM KCl, 2 mM $MgCl_2$, 10 μM $ZnSO_4$, 10 mM 2 mercaptoethanol, 0.1% NP-40, 10% glycerol, 0.2 mM EDTA and 5 μg/ml sheared salmon sperm DNA. Polyclonal rabbit α-GATA2 (H-116) antibody (Cat. No. sc-9008; Santa Cruz Biotechnology, Inc) (1:100) was added to nuclear lysates for 20 min prior to addition of probe to demonstrate GATA2 as the binding protein. To assess the specificity of the binding, 200-fold excess of each unlabeled probe was used as competitor. The mixtures were resolved in 6% non-denaturing polyacrylamide gels made in 0.5x TGE buffer (12.5 mM Tris-HCl, pH 8.5, 85 mM glycine and 0.5 mM EDTA) and the electrophoresis was performed at 4°C. For EMSA-Western blots, the experiment was carried as described above, except that the shifted DNA oligonucleotides-protein complexes were transferred onto nitrocellulose membrane, instead of PVDF. The membrane was probed with monoclonal mouse α-GATA2 (CG2-96) antibody (Cat. No. sc-267; Santa Cruz Biotechnology, Inc) and detection was performed as mentioned above.

Determination of genes differentially expressed in the presence of GATA2 mutants

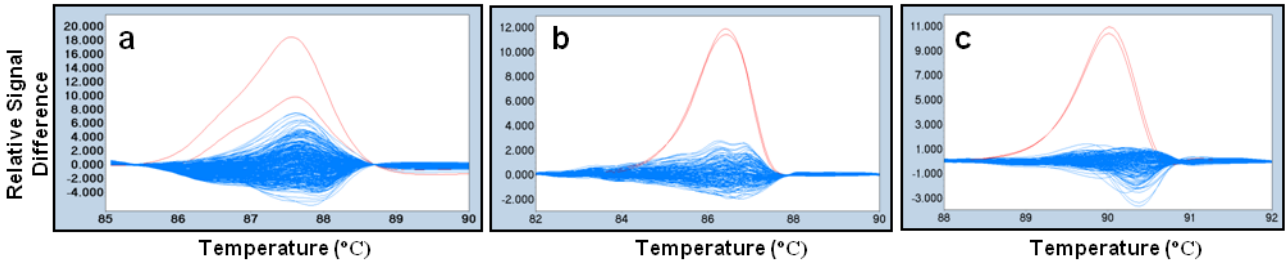
HL-60 cell lines were treated with 100 nM 4HT to turn on GATA2 WT and mutant protein expression. After 24 h, gene expression levels were determined by microarray (**Supplementary Note**).

A-3 SUPPLEMENTARY INFORMATION 1

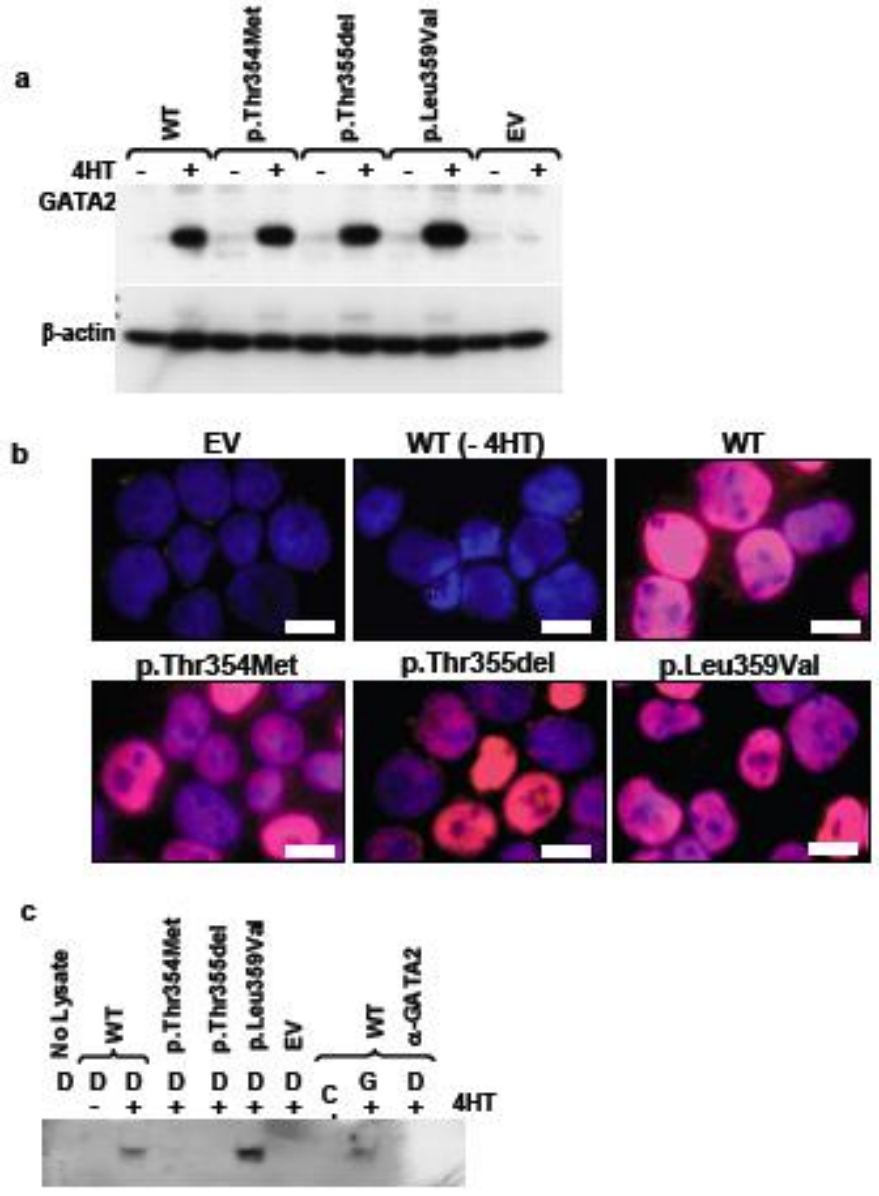
Supplementary Figure 1. Identification of novel germline p.Thr354Met and p.Thr355del variants in the highly conserved zinc finger 2 domain of GATA2 that is associated with MDS-AML. Sequence analysis was performed on representative AML or MDS affected individuals from the four pedigrees in **Fig. 1**. **a.** p.Thr354Met and p.Thr355del germline sequence variants. **a** c.1061C>T heterozygous novel variation is seen in Pedigree 1, but not in wildtype (WT) sequence. An identical variation occurs in Pedigrees 2 and 3. A heterozygous c.1063-1065delACA occurs in Pedigree 4 (blue). **b.** p.Thr354 and p.Thr355 in GATA2 from different species and the corresponding Thr residues in other GATA proteins are highly conserved. These first 2 of 5 consecutive Thr residues within the zinc finger 2 domain are totally conserved in species from Human to Zebrafish and in all six GATA binding proteins.



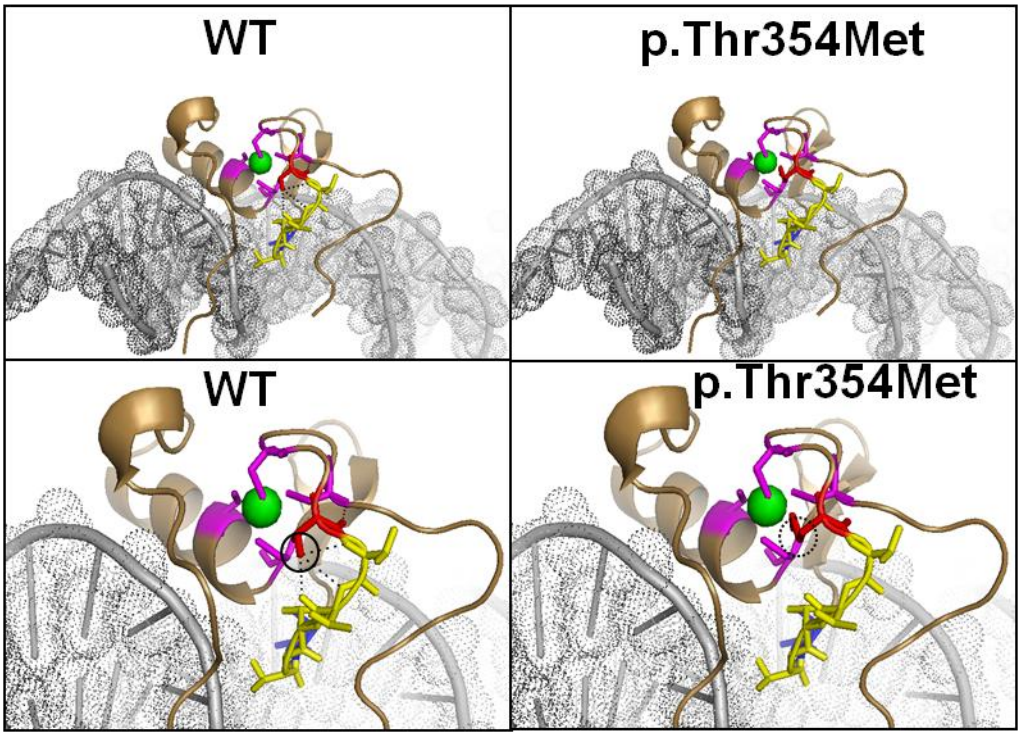
Supplementary Figure 2. Genetic analysis of *GATA2* gene variation in familial and sporadic samples. To screen for *GATA2* coding variants, high resolution melt across coding exons was performed on genomic DNA from human blood or bone marrow samples. **a.** Screen of *GATA2* exon 5 in 344 normal individuals; p.Thr354Met (red; Pedigree 2, III-4 and 3, IV-1) and wildtype p.Thr354 (blue). **b.** Screen of *GATA2* exon 5 in 268 AML patient samples; Pedigree 1, IV-1 with p.Thr354Met (red, duplicate) and wildtype p.Thr354 (blue). **c**. Screen of *GATA2* exon 2 in 268 AML patient samples; a unique variant c.182C>T (p.Ala61Val) was detected in one sample (red, duplicate).



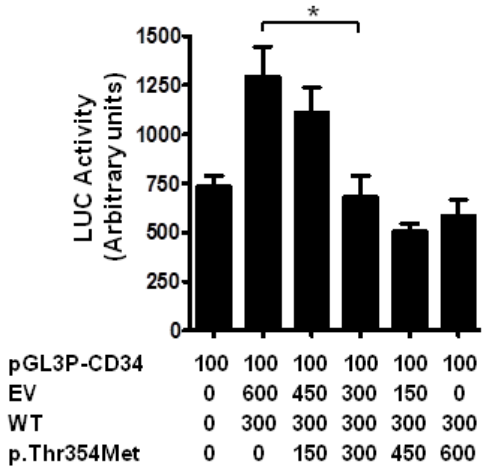
Supplementary Figure 3. Subcellular localisation and DNA binding properties of GATA2 WT and mutants in HL-60 cells. HL-60 cells carrying stably transduced 4HT-regulatable GATA2 (WT, p.Thr354Met and p.Leu359Val) or empty vector were treated with and without 4HT for 24 h. **a.** Western blot analysis of regulated GATA2 expression. Western blots of total cell lysate were probed for GATA2 and β-actin expression. **b.** Subcellular localisation of GATA2 WT and mutants. Cells from (**a**) were stained for GATA2 (red) and DAPI (blue). Scale bars, 5 µm. **c.** Electromobility shift assay (EMSA)-western blot of GATA2 WT and mutants. Nuclear lysates were prepared from cells in (**a**) and bound to GATA binding site oligonucleotides from the human TCRδ enhancer (**D**), GATA consensus (**C**) and GM-CSF promoter (**G**). The probes were visualised using chemiluminescence. A neutralizing α-GATA2 antibody in the far right lane specifically removes GATA2.



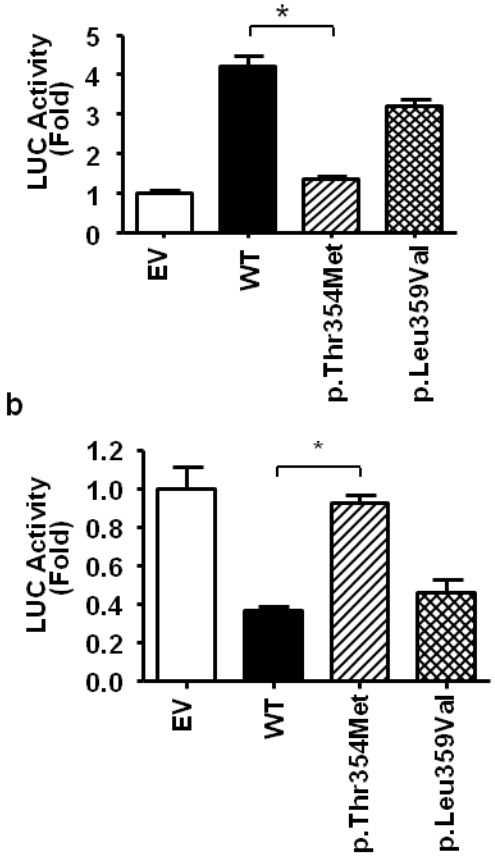
Supplementary Figure 4. Predicted structure of GATA2 zinc finger 2 binding to DNA. Human GATA2 ZF2 shares 97% sequence identity with human and mouse GATA3. We therefore used the known murine GATA3 ZF2 structure bound to DNA [7] to evaluate the effect of the observed mutations (**Supplementary Note**). The four zinc-coordinating cysteine residues are shown in purple with the zinc atom in green. Replacement of the threonine (red) at amino acid 354 in WT GATA2 with a methionine (red) disrupts interactions (circle and dotted lines) between adjacent threonines (yellow) and cysteine 349 (purple). This also introduces a bulky side-chain (dotted circle) that is predicted to alter the ZF structure or stability p. Leucine 359 (blue; residue for p.Leu359Val mutation in CML BC) projects into the major groove and contacts DNA)



Supplementary Figure 5. p.Thr354Met acts as a dominant negative mutation over WT GATA2 on the *CD34* enhancer element. HEK293 cells were cotransfected with consistent amounts of 1) *CD34* enhancer-LUC reporter (100 ng) and 2) WT (300 ng), with increasing amounts of 3) p.Thr354Met (0 - 600 ng). EV was added appropriately to control for promoter squelching and transfection efficiency. After 20 h, cells were harvested and luciferase assays performed (*p<0.05). A dominant negative activity of p.Thr354Met over WT became visible at a WT:p.Thr354Met mole ratio of 2:1, but became significant at and beyond a ratio of 1:1.

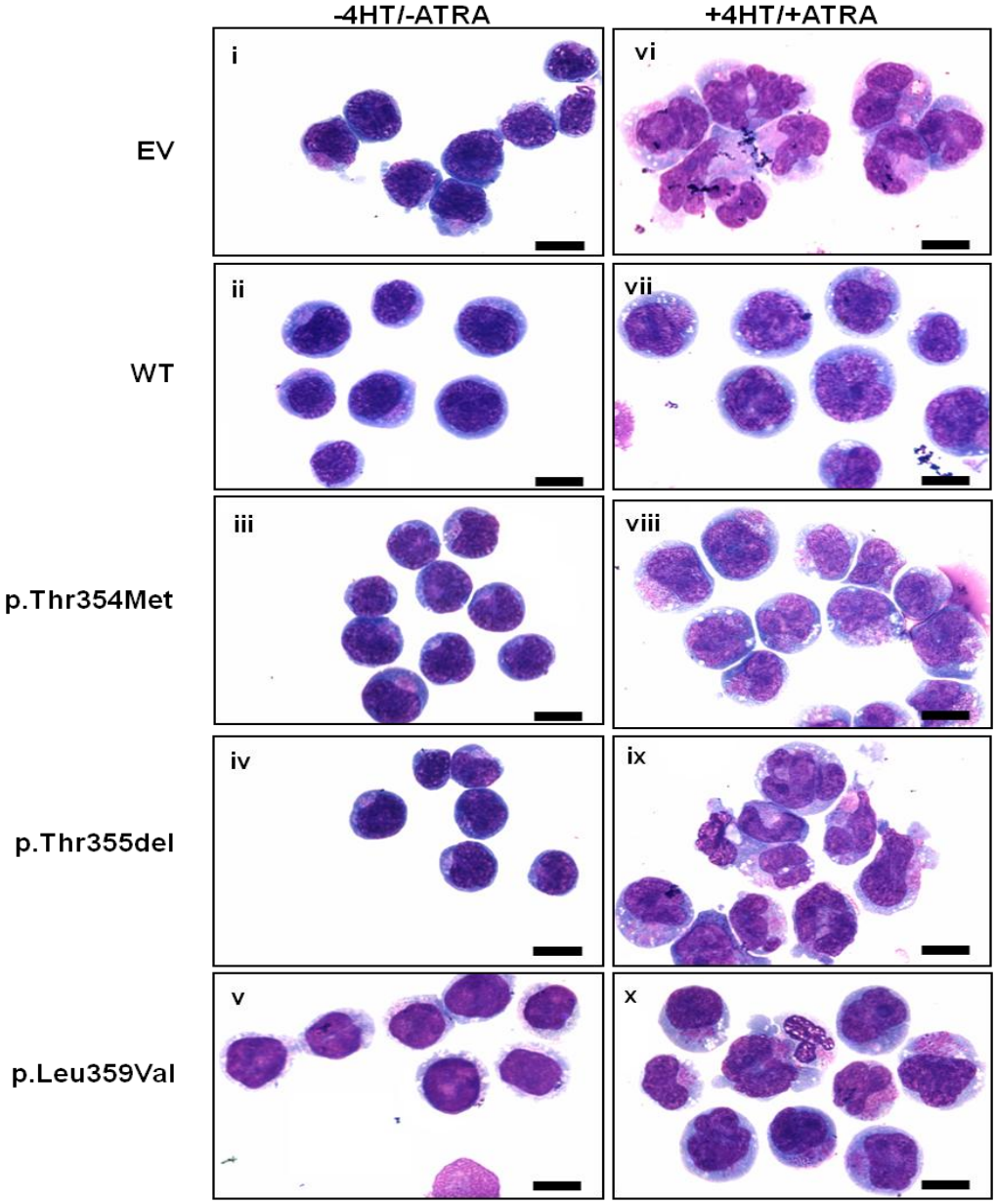


Supplementary Figure 6. p.Thr354Met displays loss-of-function, despite opposite responses on *RUNX1* enhancer in different cell types. HEK293 and Cos-7 cells were transiently co-transfected with EV, WT, p.Thr354Met or p.Leu359Val and the *RUNX1* enhancer element linked to a LUC reporter. After 20 h, cells were harvested and luciferase assays performed and plotted as fold (mean ± SEM) compared to EV control. p.Thr354Met displays loss-of-function compared to WT in both cell types (*p<0.05).

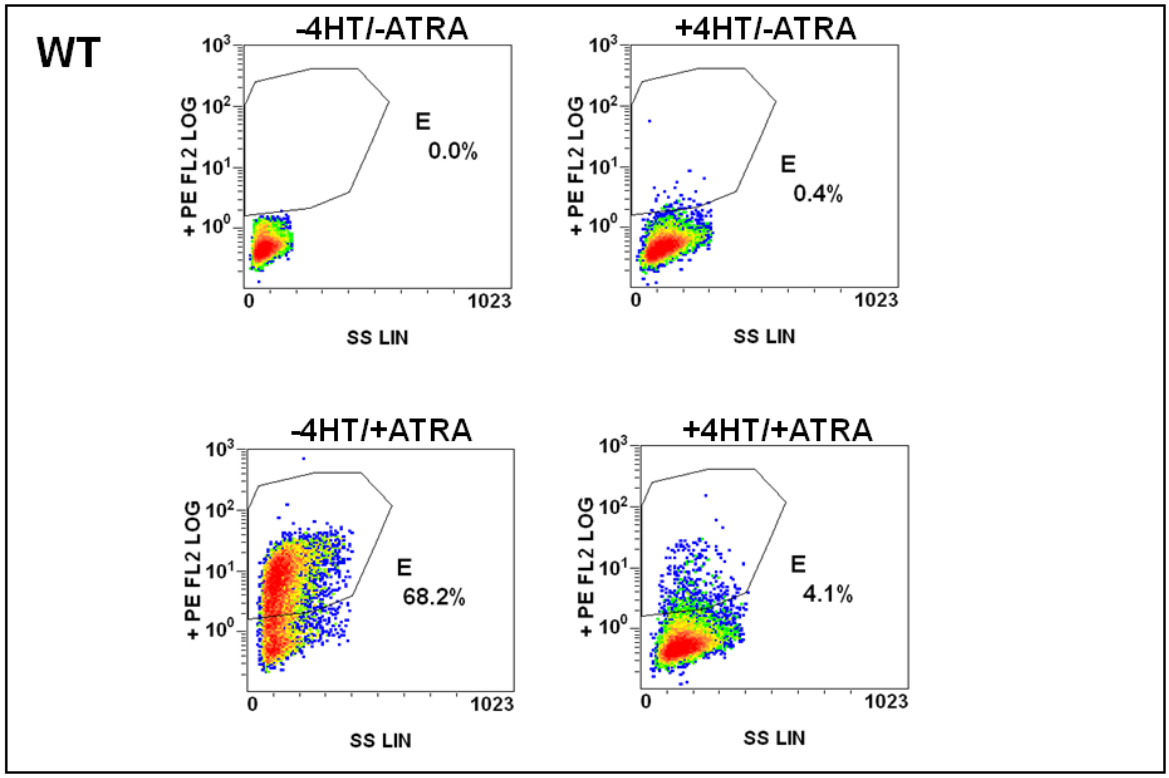


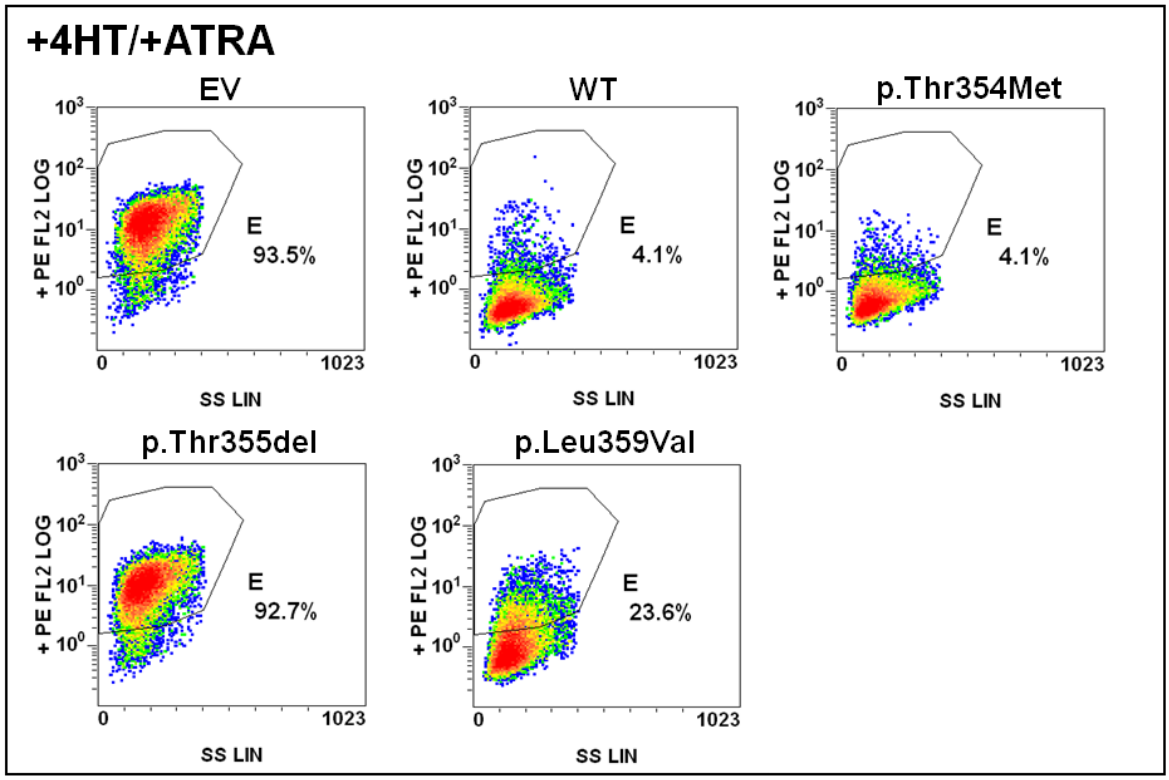
Supplementary Figure 7. p.Thr354Met inhibits ATRA-induced differentiation. HL-60 cells carrying stably transduced 4HT-regulatable GATA2 (WT, p.Thr354Met, p.Thr355del and p.Leu359Val) or EV were treated with or without 30 nM 4HT for 24 h and then with or without 2 µM ATRA for 6 days. **a.** Morphological differentiation of HL-60 cells into granulocytes. HL-60 cells were hematoxylin and eosin stained and photographs taken. Scale bars, 5 µm. **b.** CD11b expression following differentiation with ATRA. The level of surface CD11b expression was measured by FACS analysis using Phycoerythrin (PE) anti-CD11b. CD11b positive cells reside in region E. Indicative FACS plots relating to GATA2 WT (top panel – from Figure 6a-d) and +4HT/+ATRA (bottom panel from Figure 6d) are shown. **c.** Apoptosis following differentiation with ATRA. Cells were FACS analysed following staining with FITC anti-Annexin V and propidium iodide (PI). Indicative FACS plots of PI vs Annexin V relating to GATA2 WT (top panel – from Figure 6i-m) and +4HT/+ATRA (bottom panel from Figure 6m) are shown.

Supplementary Figure 7a. Morphological differentiation of HL-60 cells into granulocytes.

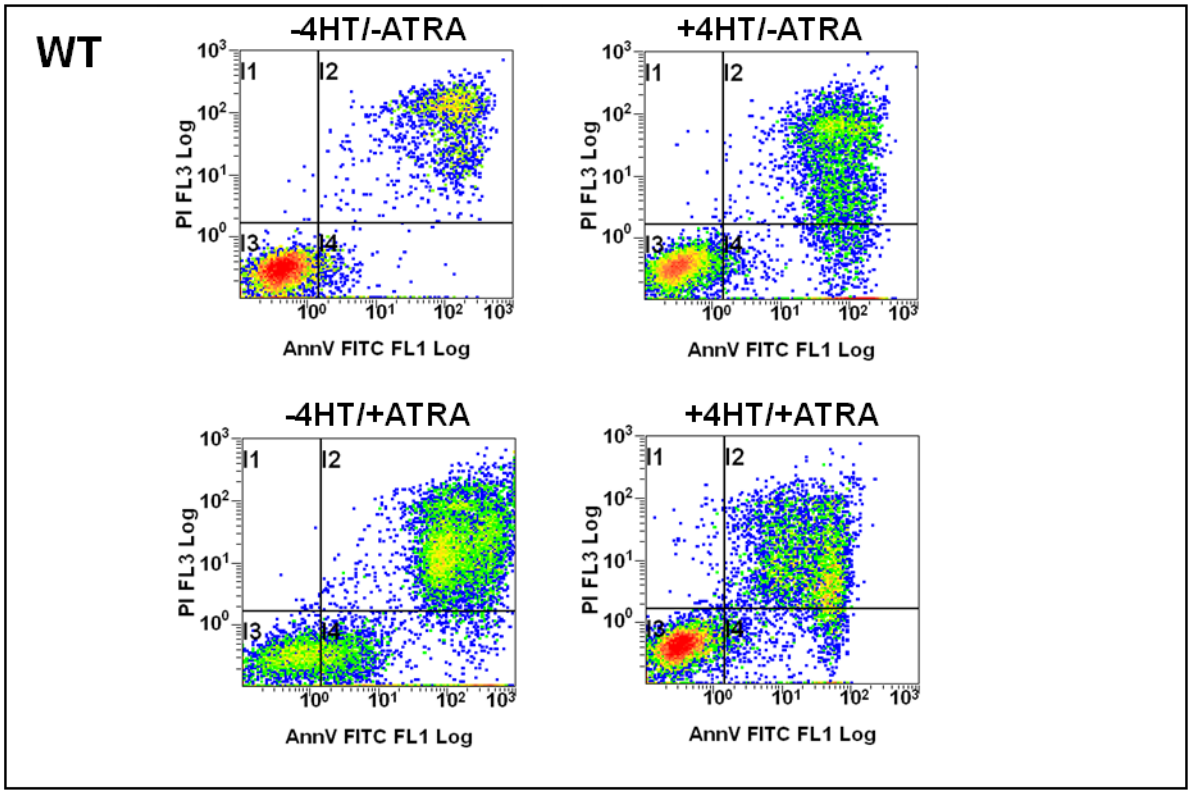


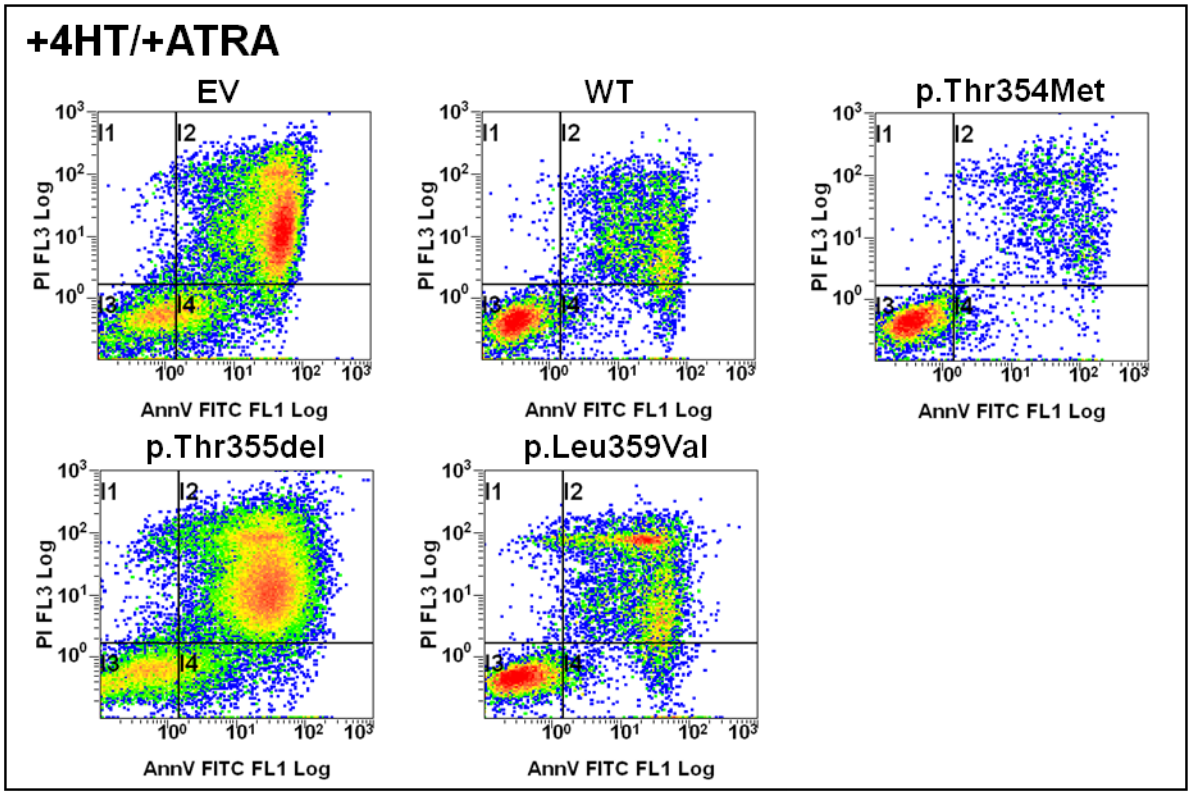
Supplementary Figure 7b. CD11b expression following differentiation with ATRA.
Density plots – CD11b vs Side Scatter.





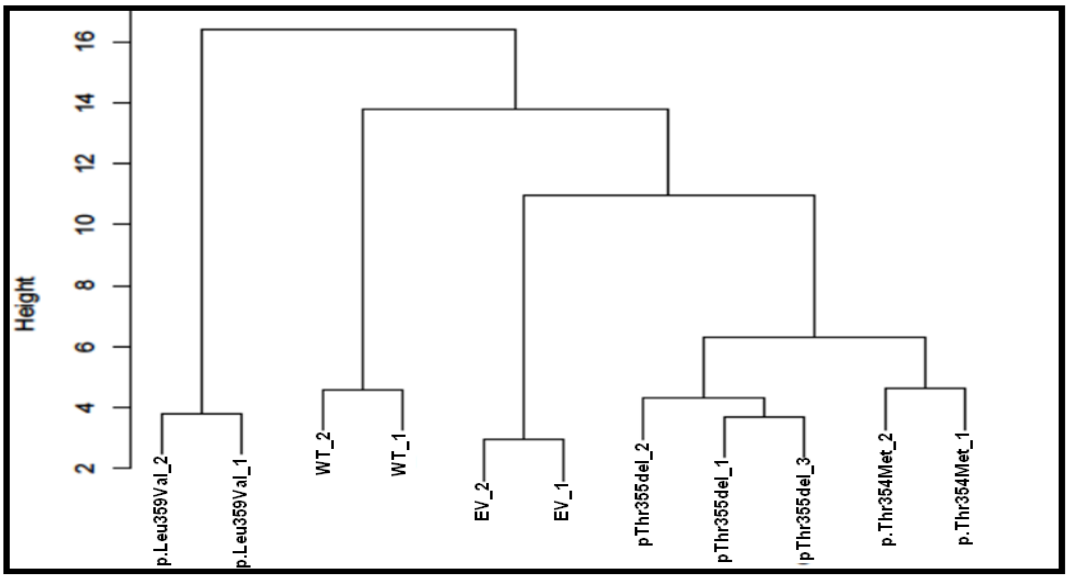
Supplementary Figure 7c. Apoptosis following differentiation with ATRA. Density plots – Propidium iodide vs Annexin V. Annexin V^+, PI^- (region I4) (black) or Annexin V^+, PI^+ (region I2) (white).



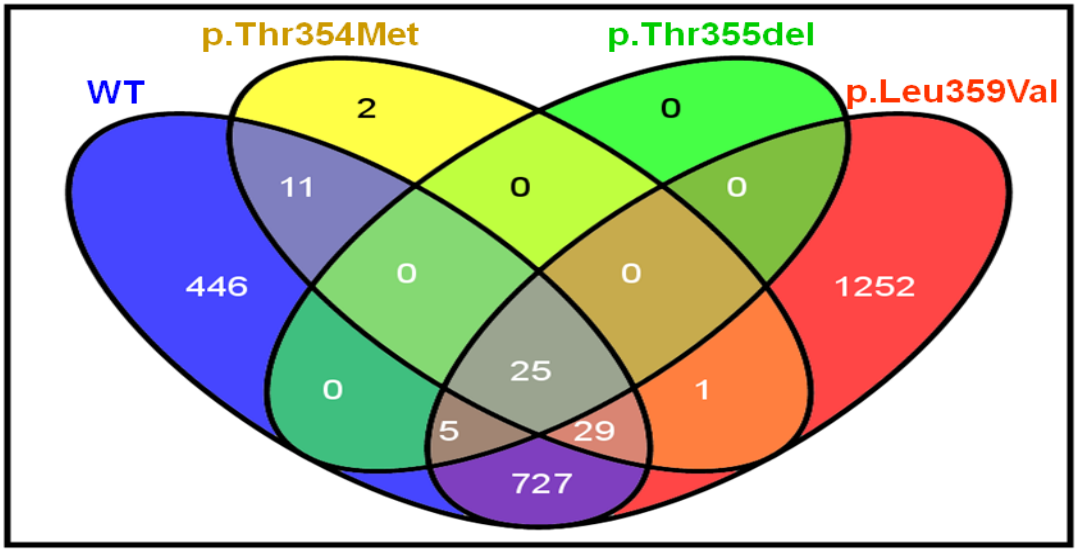


Supplementary Figure 8. Differentially regulated genes upon expression of GATA2 WT and mutants in HL-60 cells. HL-60 cell lines were treated with 4HT to turn on GATA2 WT and mutant protein expression. After 24 h, gene expression levels were determined between EV and GATA2 expressing cell lines (**see Supplementary Table 6**). **a.** A cluster dendrogram shows GATA2 mutants p.Thr354Met and p.Thr355del being most similar to each other and most closely related to the EV suggesting loss-of-function (LOF). It also shows that p.Leu359Val is quite different to these LOF mutants demonstrating rather gain-of-function compared to WT. **b.** Differentially regulated genes are shown in a 4 Way Venn Diagram. **c.** The number of genes in common between any two GATA2 proteins is tabulated.

a



b



c

	WT	p.Thr354Met	p.Thr355del	p.Leu359Val
WT	1,243	65	30	786
p.Thr354Met		68	25	55
p.Thr355del			30	30
p.Leu359Val				2,039

Supplementary Table 1. Panel of candidate oncogenic genes in AML. Primers were designed and generated to amplify all coding exons; some exons required multiple amplicons. Primer sequences are available upon request. The sequence of the resultant amplicons was determined by DNA sequencing and this was compared to the RefSeq database to identify variants/mutations. Genes were chosen on the basis of known or predicted role in the network of hematopoietic regulators based on the following criteria: known to be mutated in familial myeloid leukemia, displaying hematopoietic expression pattern(s), a known role in hematopoiesis, a candidate myeloid transcription factor [1], mutated in mouse models with hematopoietic defects, known somatic mutations in myeloid leukemia, *RUNX1* target genes or cofactors, common retroviral insertion sites, or members of the ETS transcription factor family.

Gene	Exons (coding)	Amplicons
ATM	63	72
BMI1	10 (9)	9
CALR	9	9
CCND3	5	5
CEBPA	1	4
CEBPB	1	3
CEBPD	2	4
CEBPE	1	3
KIT	21	21
CSF1R	22 (21)	22
CSF3R	17 (15)	16
ELF1	9 (8)	11
ELF2	9 (8)	9
ERG	11 (9)	12
ETS1	8	8
ETS2	10 (9)	9
ETV7	8	8
FIP1L1	18	18
FLI1	11 (10)	30
FLT3	24	24
ZFPM1	10	11
GATA2	6 (5)	7
GATA3	6 (5)	6
ID2	5 (2)	2
JAK2	25 (23)	23
JAK3	24 (23)	23
KLF3	6 (5)	6
KLF4	5 (4)	6
KLF5	4	7
KLF11	4	5
KRAS2	6 (4)	4
LEF1	12 (11)	11
MASTL	12	15
MEIS1	13 (12)	12
MLF1	7	7
MLL	36	48
MYB	15	14
NBN	16	16
NPM1	12 (11)	10
NRAS	7 (4)	4
PDGFRA	23 (22)	21
PIM1	6	6
PIM2	6	5
PTPN11	16 (15)	14
SPI1/PU1	5	5
RAPGEF1	24	27
RUNX1	11	11
SMAD3	9	9
SOX4	1	3
TAL1	4 (3)	3
Total Amplicons		638

Supplementary Table 2. Clinical details with genotype of individuals in four families harboring the MDS/AML-predisposing p.Thr354Met and p.Thr355del mutations.
Asterisk - FAB classification of AML if known; Leukemia - refers to non-specific leukemia; unknown - disease status unknown, BMT – bone marrow transplant. FAB classification M2, M7; refractory anemia (RA); refractory cytopenia with multilineage dysplasia (RCMD); coronary artery disease (CAD); cardiovascular accident (CVA); Cytogenetically normal (CN).

Pedigree	Individual	Disease*	GATA2 Genotype	Age at diagnosis (years)	Age of Death (years)	Karyotype
1	I-1					
	I-2	**Leukemia**			42	
	II-1				alive	
	II-2	**MDS-AML**		36	40	
	II-3				41	
	II-4	**AML**			50	
	II-5	**AML** (and ALL)		10	10	
	II-6				68	
	II-7				62	
	II-8				78	
	II-9	**AML** (and CML?)		27	29	
	III-1		p.Thr354		alive	
	III-2	**MDS-AML**		16	34	1q abnormality (isochromosome)
	III-3		p.Thr354		alive	
	III-4	**MDS-AML**	**p.Thr354Met**		32	
	III-5		**p.Thr354Met**		alive	
	III-6				54	
	III-7		p.Thr354		alive	
	III-8		**p.Thr354Met**		alive	
	III-9	**MDS-AML**	**p.Thr354Met**		30	
	III-10	**AML M2**		13	14	
	III-11				alive	
	III-12	**AML M2**		19	19	
	III-13	**Leukemia (myeloid?)** , pancytopenia		18	19	
	IV-1	**MDS-AML**(received BMT)	**p.Thr354Met**	teenager	alive	-7, +21
	IV-2	**MDS-AML**	**p.Thr354Met**	15	15	
	IV-3		p.Thr354		alive	
	IV-4		p.Thr354		alive	
	IV-5				alive	
	IV-6				alive	
	IV-7				alive	
	IV-8				alive	
2	I-1				alive	
	I-2	**Acute Leukemia** , anemia, leukopenia, splenomegally			46	
	I-3	unknown			8 siblings	
	II-1				alive	
	II-2	**AML**			22	
	II-3				alive	
	II-4	**MDS**;received BMT	**p.Thr354Met**	45	alive	-7, +8
	II-5	**Atypical CML** (eosinophilia), pancytopenia			24	
	II-6		**p.Thr354Met**		alive	
	II-7				alive	
	III-1		p.Thr354		alive	
	III-2		p.Thr354		alive	
	III-3				alive	
	III-4	**MDS**	**p.Thr354Met**	20	alive	-7
	III-5		p.Thr354		alive	
	III-6				alive	
	IV-1				alive	

Pedigree	Individual	Disease*	GATA2 Genotype	Age at diagnosis (years)	Age of Death (years)	Karyotype
3	I-1	cerebral haemorrhage			49	
	I-2				85	
	II-1	pernicious anemia & congestive heart failure			72	
	II-2	**Leukemia** ?	**p.Thr354Met**		59	
	II-3	**AML M2**			41	
	II-4	macrocytic anemia & cerebral haemorrhage	**p.Thr354Met**		82	
	II-5				alive	
	II-6				alive	
	II-7	haemorrhagic stroke & breast			62	
	II-8	**Leukemia**		44	44	
	II-9	**AML M2**	**p.Thr354Met**	32	32	const. 9p21-22 variation
	III-1	**MDS-AML M2**	**p.Thr354Met**	38	42	CN
	III-2				alive (~74)	
	III-3				alive (~66)	
	III-4	**MDS-AML**	**p.Thr354Met**	23	28	
	III-5	unknown			2 siblings	
	III-6	**MDS(RA)-AML**	**p.Thr354Met**	35	40	
	III-7	unknown			2 siblings	
	III-8				alive	
	III-9	well	**p.Thr354Met**		alive (~61)	
	III-10	unknown			3 siblings	
	III-11	unknown			2 siblings	
	III-12				alive	
	III-13	haemorrhagic stroke			unknown	
	III-14	unknown			4 siblings	
	III-15	unknown			4 siblings	
	IV-1	**AML M7** ; BMT at 21 years	**p.Thr354Met**	21	alive (~39)	-7, +8, t(1q;7p)
	IV-2	unknown			2 siblings	
	IV-3	Few abnormal BM cells (**MDS**?); BM donor for IV-1		16	alive (~33)	CN
	IV-4	well			alive	
	IV-5	**MDS**	**p.Thr354Met**	15	alive (32)	-7, const. 9p21-22 variation
	IV-6	well			alive	
	V-1	well			alive	
4	I-1	CAD; died CVA			unknown	
	I-2	died cirrhosis			unknown	
	II-2	**MDS** (RCMD); BMT at 53 years	**p.Thr355del**	53		+8 (2/20 cells)
	III-1	**MDS**; BMT at 15 years	**p.Thr355del**	13		-7

Supplementary Table 3. **Clinical summary of affected individuals from Pedigree 1.** Patient ID relates to Pedigree 1 (Fig. 1a). Age of death (Age) and diagnosis were verified with "Births, Deaths and Marriages Registries - Australia" or pathology reports. Asterisk – still alive.

Patient ID	Diagnosis	Age	Verification
I-2	Leukemia	42	Death certificate
II-2	Leukemia	40	Death certificate
II-4	AML	50	Death certificate
II-5	Acute lymphocytic leukemia	10	Death certificate
II-9	CML	29	Death certificate
III-2	MDS to AML	38	Pathology report
III-4	AML	32	Pathology report
III-9	MDS to AML	32	Pathology report
III-10	AML	14	Death certificate
III-12	AML	19	Death certificate
III-13	Pancytopenia	19	Death certificate
IV-1	MDS (BMT at 26 years in 2003)	33*	Pathology report
IV-2	MDS to AML	15	Pathology report

Supplementary Table 4. Cell lines screened for *GATA2* mutations.

Name	Origin	Cell Type
Ball1	B-ALL	B-Lymphoblastic
Daudi	Burkitt's Lymphoma	B-Lymphobalstic
HL-60	AML M2	Promyeloblastic
JURKAT	T-ALL	T-Lymphocytic
K562	CML BC	pre-Erythrocytic/Megakaryocytic blasts
KG-1a	AML	Promyeloblastic
KU812	CML BC	Basophilic blasts
REH	ALL (non-T; non-B)	pre-B-Lymphoblastic
U937	Histiocytic Lymphoma	Monoblastic
KCL-22	CML BC	Myeloblastic
KT-1	CML BC	Myeloblastic
KYO-1	CML BC	pre-Erythrocytic/Monocytic blasts
EM-2	CML BC	Myeloblastic
LAMA-84	CML BC	Erythrocytic/Megakaryocytic blasts
MEG-01	CML BC	Megakaryoblastic
HEK293T	Embryonic kidney	Fibroblastic

Supplementary Table 5. Haplotype analysis of p.Thr354Met mutation in three families. The haplotype of genomic DNA surrounding the p.Thr354Met mutation was determined by PCR sequencing 9 selected amplicons containing a total of 50 SNPs. Shown is the name, position and genotype of 8 informative SNPs in each of the families that segregates with the c.1061C>T (p.Thr354Met) mutation. Pedigrees 1 and 2 contained a common haplotype (red) while Pedigree 3 differs (blue). *a*. Population Diversity; RefSeq SNP nucleotide and range of frequency in dbSNP in January, 2011. Asterisk; only p.Thr354Met carrying individuals in Pedigrees 1 and 2 contain these rare variations. A matched normal population screen showed 10/238 individuals heterozygous for the uncommon "T" variant of rs45463801 in GATA2 3'UTR indicating possible linkage disequilibrium of this specific haplotype. Notably, none of these contained a p.Thr354Met mutation.

Variants	Position	Population Diversity[a]	Pedigree 1	Pedigree 2	Pedigree 3
rs7629791	-17,667	36-76% C	T	C/T	C
novel	-11,748 (-460 promoter)	100% GTCT	GTCT/ΔGTCT*	GTCT/ΔGTCT*	GTCT
p.Thr354Met	**+1 (Exon 5)**	**100% C**	**T**	**T**	**T**
rs45463801	+1,429 (Exon 6 - 3'UTR)	96-98% C	C/T*	C/T*	C
rs2713590	+7,949	76-86% T	T	T	G
rs10934854	+8,092	40-74% G	G	G	T
rs6808807	+8,969	75-86% G	G	G	A
rs2135644	+9,364	50-85% C	C	C	T
rs2135645	+9,375	69-88% C	C	C	T

Supplementary Table 6. Microarray analysis to identify differentially regulated genes following expression of GATA2 WT and mutants in HL-60. HL-60 cell lines were treated with 100nM 4HT to turn on GATA2 WT and mutant protein expression. After 24 h, gene expression levels were determined between EV and GATA2 WT and mutant expressing cell lines (**See also Supplementary Figure 8**). Details of the microarray analysis available as separate online supplementary information.

Original analysis file is accessible at the publisher's website http://www.nature.com/ng/journal/v43/n10/extref/ng.913-S1.pdf **and is also provided in the CD presented at the back of the thesis.**

Supplementary Table 7. GATA2 and other hematopoietic factors: interactions and targets.

GATA2 interacting partners	Supplementary Reference
GATA1	[8]
SPI1 (PU.1)	[9]
ZFPM1 (FOG1)	[10]
ZFPM2 (FOG2)	[11]
LMO2	[12]
PML	[2]
ZBTB16 (PLZF)	[13]
RARA (RARα)	[14]
FLI1	[15]
TAL1	[12, 15]
GATA2 target genes	
GATA1	[16]
RUNX1	[17, 18]
CEBPA	[19]
FLI1	[15]
TAL1	[15]
SPI1 (PU.1)	[19, 20]
Regulators of GATA2 transcription	
GATA1	[21]
FLI1	[15]
TAL1	[15]
CEBPA	[18, 22]
HOXA9	[23]
EVI1	[24]
NOTCH1	[25]

Supplementary Table 8. Primer sequences for HRM analysis of *GATA2* coding exons. Primer sequences were designed using Primer3 [26] to span all coding exons and splice junctions. The optimal PCR amplification conditions including $MgCl_2$ concentration and touchdown PCR annealing temperatures are given. Touchdown was performed at 1ºC/cycle until the lower target temperature was achieved, and subsequent annealing was performed at this temperature to a total of 50 cycles.

Primer Name	Primer Sequence	$MgCl_2$ (mM)	Touchdown PCR
GATA2Ex2a(f)	ACCTCGTGGTGGGACTTTGG	2.0	68ºC → 65ºC
GATA2Ex2a(r)	ACCTCGTCTGGAGGCAGCAG		
GATA2Ex2b(f)	CACCCGGCCGTGCTGAAT	2.0	68ºC → 65ºC
GATA2Ex2b(r)	CGCCTGGGTTCTCATCACCA		
GATA2Ex3a(f)	TGATCTTTCTGCCCACCCTGAT	2.0	68ºC → 65ºC
GATA2Ex3a(r)	TACACAGAGAGTGGGCCTCCAG		
GATA2Ex3b(f)	GTGAGCCCCTTCTCCAAGACG	2.0	68ºC → 65ºC
GATA2Ex3b(r)	CGGAAGATGAGGCTGGAGACG		
GATA2Ex3c(f)	CCACCCAAAGAAGTGTCTCCTGA	3.5	68ºC → 63ºC
GATA2Ex3c(r)	GCCGGCACATAGGAGGGGTAG		
GATA2Ex3d(f)	CTGCGCCCAGGCCTAGCTACTAT	3.5	68ºC → 63ºC
GATA2Ex3d(r)	AAAAACGCAAATGCTCCCCTCTT		
GATA2Ex4(f)	GTTAAGCAGGCCCCCGTGT	2.5	68ºC → 63ºC
GATA2Ex4(r)	CCTGTAATTAACCGCCAGCTCCT		
GATA2Exon5(f)	CCTGCTGACGCTGCCTTG	2.5	68ºC → 62ºC
GATA2Exon5(r)	GGTCCCCGTTGGCGTTTC		
GATA2Ex5(f)	TTTAGCCCTCCTTGACTGAGC	2.0	68ºC → 60ºC
GATA2Ex5(r)	CAAGCCAAGCTGGATATTGTG		
GATA2Ex6a(f)	AGGAAGGAACTGGCCCTCTGA	2.0	68ºC → 62ºC
GATA2Ex6a(r)	TGTCCGGAGTGGCTGAAGG		
GATA2Ex6b(f)	GGAGAAGTCATCCCCCTTCAGTG	2.0	68ºC → 61ºC
GATA2Ex6b(r)	CGGTCCTCGACGTCCATCTGTT		
GATA2ΔGTCT(f)	GGCCATGTGTGTCGGTGT	2.0	68ºC → 60ºC
GATA2ΔGTCT(r)	TTCCAGGAGTGCTCACACG		

Supplementary Table 9. Primers for GATA2-responsive promoter or enhancer-luciferase constructs. Underlined are the *Kpn*I, *Bgl*II and *Sfi*I restriction endonuclease sites used for cloning.

Primer Name	Primer Sequence
RUNX1enh-F	GGGGTACCCAGTTTCATGTGGGGGTAGG
RUNX1enh-R	GAAGATCTCTCGTAAATTTTGGCCTCCT
FLI1enh-F	GGGGTACCTCCTCCTCGAAATCTTGCTC
FLI1enh-R	GAAGATCTTAGCTCATCGACCCCTGGTA
LYL1prom-F	CTCTGGCCTAACTGGCCTATGTGATCCTGTAGCCAAGAG
LYL1prom-R	CTCTGGCCGCCGAGGCCCCCAGCACTGTTCTTGCAGCCT
CSF1Rprom-F	CTCTGGCCTAACTGGCCAGATATGCATTACTTTGGAGATTCCAAGG
CSF1Rprom-R	CTCTGGCCGCCGAGGCCCCTCGGTGGGGAAGTGGCA

Supplementary Table 10. Haplotype mapping primers pairs. Primer pairs were designed to amplify regions flanking the site of the GATA2 p.Thr354Met variant. Following PCR amplification, these amplicons were DNA sequenced and the SNPs identified for haplotype analysis.

Primer Name	5' Primer Sequence	3' Primer Sequence
GATA2-SNP1	TGGTTTACAGTACCTGGAAGGAGAAAG	TCCTGGAAGAGGAGCTGTGTGAC
GATA2-SNP3	AAACACAACGTAGGGTGGGCACT	AATGCATCATTTACCACCGCAGAT
GATA2-SNP5	TTGGAGAAAGACCTATGGACAGCAG	TCAGAAAGCTTGGGACACGTCTTTA
GATA2-SNP7	TTAGGTTACAAAGGGAGGGCAAACT	CACCATTTGACCTGGGTGCTTC
GATA2-Exon5	CCTGCTGACGCTGCCTTG	GGTCCCCGTTGGCGTTTC
GATA2-SNP13	ACAGACGAAGGCAACCATTTTTAGA	GCTCCTGTGGCTACGTACAATCAAC
GATA2-SNP15	CTGATCACCAATATTTCATCCATGC	GCAGAGACTTTGAAGACGATTCAGC
GATA2-SNP17	CTGGAGTGCAGACGTGCTATCAT	GGCCACAAAATTCAAACTGTTTACC
GATA2-SNP18	GTCGGCCTTCTTCCCTCTGTTT	AAATAATAGGGTCTGTTGCGGGTTT
GATA2-SNP19	ACAAAAACCTGTCTTCTCTGCCTTG	CTTCTAAGATTGGAGGTGGGTTTGG
GATA2-SNP20	CTCTCCTTTGTGCCAGCCTCCT	CTCTCCTAACTCCTGGAGCTTGTGC

Supplementary Table 11. Primers for GATA2 mutagenesis. Nucleotide changes leading to p.Thr354Met and p.Leu359Val mutations (bold underlined). We also introduced a single silent change (underlined) that generated a *Mfe*I site in p.Thr354Met clones and destroyed a *Nar*I site in p.Leu359Val clones.

Primer Name	Primer Sequence
T354M-F	GCACCTGTTGTGCCAATTGTCAGA**T**GACAACCACCACCTT
T354M-R	AAGGTGGTGGTTGTC**A**TCTGACAATTGGCACAACAGGTGC
L359V-F	ATTGTCAGACGACAACCACCACC**G**TATGGCGACGAAACGCCAACGGG GACCCTGTCT
L359V-R	AGACAGGGTCCCCGTTGGCGTTTCGTCGCCATA**C**GGTGGTGGTTGTC GTCTGACAAT
355delT-F	ATTGTCAGACG|ACCACCACCTTATGGCGCCGAAACGCCAACGGGG
355delT-R	CCCCGTTGGCGTTTCGGCGCCATAAGGTGGTGGT|CGTCTGACAAT

Supplementary Table 12. Primers for generation of *GATA2* and *GATA2*-FLAG tagged cDNAs. Underlined are the *Xba*I restriction endonuclease sites used for cloning.

Primer Name	Primer Sequence
KOZAK-GATA2-F	GTTCTAGAGCCACCATGGAGGTGGCGCCGGA
GATA2-FLAG-R	GTTCTAGACTACTTGTCATCGTCGTCCTTGTAGTCGCCCATGGCGGTC ACCATGC
GATA2-R	GTTCTAGACTAGCCCATGGCGGTCACCATGC

Supplementary Table 13. Primers for electromobility shift assay (EMSA). GATA binding sites (underlined).

Primer Name	Primer Sequence
Human TCRδ enhancer-F	CACTTGATAACAGAAAGTGATAACTCT
Human TCRδ enhancer-R	AGAGTTATCACTTTCTGTTATCAAGTG
GATA Consensus-F	CAGGGACATGATAAGGGAGCCAA
GATA Consensus-R	TTGGCTCCCTTATCATGTCCCTG
Human GM-CSF-153 promoter-F	TCTCTCGTGATAAGGATCCTGGA
Human GM-CSF-153 promoter-R	TCCAGGATCCTTATCACGAGAGA

A-4 SUPPLEMENTARY NOTE

Clinical information

Pedigree 1

Pedigree 1 with MDS transforming to AML was described in 1983 [27]. In 1983 there were six individuals who had been diagnosed with leukemia (one with a prior MDS phase) and three with MDS only. A total of 13 family members have now been diagnosed with leukaemia or related disorders (**Fig. 1a**). The diagnoses are summarized in **Supplementary Table 2** and **3**.

The proband (case 18 [27]; III-2) had an underlying psoriasis that began at age 18 years. He also presented in 1978 with a five month history of suppurating lesions consisting of 1 cm diameter ulcers with a purulent discharge. These ulcers were responsive to treatment with prednisolone and were most likely due to a neutrophilic dermatosis such as Sweet's syndrome or Pyroderma gangrenosum, which are often associated with an underlying MDS or AML. Blood and bone marrow analyses at the time confirmed an MDS phenotype that later transformed into AML.

His sister (III-4) also had psoriasis and a six year history of raised erythematous lesions prior to a marrow biopsy in 1979 which revealed moderate hypocellularity and dysplastic myeloid and erythroid cells. In 1982, she presented with ulcerating lesions and her bone marrow was hypoplastic and dysplastic, suggestive of MDS. She later developed AML (after publication [27]). Neither psoriasis nor these ulcerating lesions have been documented for other family members suggesting that they may not be linked to the AML predisposition. The psoriasis may be linked to the chromosome 1q variation identified in case 18 (III-2), as this chromosome has been previously linked to this condition [28]. The ulcerating lesions are likely to be due to a neutrophilic dermatosis associated with the MDS in these two individuals, but not linked to the underlying genetic mutation responsible for the AML predisposition.

Recent analysis of the family indicates that both the son and daughter (IV-1 and IV-2) of the proband developed MDS. IV-1 underwent a bone marrow examination which identified

refractory cytopenia with multi-lineage dysplasia. His bone marrow was hypocellular with active erythropoiesis and granulopoiesis, moderate dysplasia with macronormoblastic maturation and inadequate hemoglobinization. Cell counts revealed $3.3x10^9$ cells/l white cells, $121x10^9$ cells/l platelets, scanty neutrophils and less than 5% blasts. One paratrabecular focus of mononuclear cells, which may represent myeloblasts, was identified. Karyotyping revealed monosomy 7 in all cells and trisomy 21 in a small percentage of cells. He underwent a bone marrow transplantation in 2003 at age 26 years and is currently doing well.

The mother of the proband (II-2) also developed MDS-AML, as did her nephew (III-9). Her three sisters (II-4, II-5 and II-9) all developed AML while her mother (I-2) died of unspecified leukemia. Two of the proband's cousins (III-10 and III-12) died of AML while another cousin (III-13) died of unspecified leukemia.

Pedigree 2. This previously undescribed pedigree consists of five affected individuals over three generations (Fig. 1A). Individual I-2 suffered of acute leukemia, II-2 of a not further characterized AML, II-4 and III-4 of MDS with monosomy 7 with II-4 having in addition a trisomy 8.

Individual II-5 was a previously well 24 year old man who presented with pancytopenia, peripheral eosinophilia, and recurrent pneumonia in the year preceding his death in 1971. Two separate bone marrow biopsies were notable for hypercellularity with eosinophilia. Karyotype was not performed. Skin biopsy of the scalp showed evidence of hyperkeratosis, papillomatosis, and inflammation. On his final hospital admission, left lung infiltrates were associated with progressive respiratory compromise and lung biopsy revealed chronic interstitial pneumonitis and fibrosis. A diagnosis of "Atypical CML" was made, in part, on the basis of eosinophilic lung infiltrates. The patient was treated with prednisone for a period before his respiratory status worsened and he developed hepatosplenomegaly. The patient eventually succumbed to an acute hypotensive episode leading to cardiopulmonary arrest.

Pedigree 3. This extensively described pedigree has eleven affected individuals over three generations (**Fig. 1a**) [29]. Individual IV-1 had acute megakaryoblastic leukemia (FAB M7)

and was transplanted with his brother's (IV-3) bone marrow when 21 years of age and is still alive 18 years later. IV-3 and IV-5 had MDS, III-1, III-4 and III-6 all had MDS evolving into AML, II-3, II-9 had AML, and II-8 died of a not further specified leukemia. Karyotyped individuals IV-1 and IV-5 showed monosomy 7 in the affected cells with IV-1 also displaying trisomy 8 and t(1q:7p).

High resolution melt mutation screening

High resolution melt (HRM) analysis on the LC480 Lightcycler (Roche) using a LightCycler® 480 High Resolution Melting Master (Roche) according to the manufacturer's protocol was performed to develop a sensitive screen for mutation detection. Primer pairs were designed across each exon (multiple primer sets were used to span large exons) such that amplicons were no larger than 250 bp (**Supplementary Table 8**). All positive samples containing putative variants were confirmed by DNA sequencing. As a confirmation of the assay for detection of T354M mutations, in comparison with DNA sequencing, the HRM assay produced 4% false positives and 0% false negatives (data not shown). All positives were checked by PCR sequencing.

Modeling of GATA structure

The human GATA2 C-terminal zinc finger protein sequence was uploaded to Swiss Model (an automated comparative protein modeling server). This generated a GATA2 ZF model based on PDB code:3dfx (mouse Gata3 [30]; 97% sequence identity, Fig. 1E; E-value $8.6x10^{-21}$). We used the PyMol program to visualise and predict the effects of the T354M mutation [31]. The PDB file is available upon request.

Determination of genes differentially expressed in the presence of GATA2 mutants.

HL-60 cell lines were treated with 100 nM 4HT to turn on GATA2 WT and mutant protein expression. After 24 h, gene expression levels were determined by microarray. Total RNA was purified using the RNeasy protocol (Qiagen) and converted to biotinylated sense DNA as per the Affymetrix Genechip Whole Transcript (WT) Sense Target Labeling Assay manual. Genechips were hybridized overnight at 45ºC for 17 h in an Affymetrix Hybridization oven, washed and scanned using the Affymetrix GSC3000 scanner.

Raw microarray data were normalized with robust multichip average (RMA) methods [32] using aroma.affymetrix analysis [33]. Quality control was assessed using normalized unscaled standard error (NUSE) plot as implemented in aroma.affymetrix analysis. Batch effect was removed using ComBat analysis [34]. The differential gene expression was determined with the Empirical Bayes and moderated t-statistic using linear model for microarray analysis (LIMMA) [35]. Differential gene expression was adjusted for multiple testing to control the false discovery rate (FDR) using Benjamini-Hochberg method [36]. Genes with FDR $p<0.05$ were considered statistical significant. All statistical analyses were performed using R statistical software.

A number of analysis including ChIP Enrichment Analysis, Motif Activity Response Analysis and Gene Set Enrichment Analysis (see **URL**) were performed on the these array data. Among the patterns revealed were targets of the MYC oncogene such as numerous ribosomal proteins as a predominant feature for GATA2 WT, but not for p.Thr354Met or p. Thr355del. Indeed, MYC expression in our arrays is down regulated by GATA2 WT, but not by p.Thr354Met or p.Thr355del (**Supplementary Table 6**) consistent with repression of *MYC* by GATA2 WT being abrogated in p.Thr354Met and p.Thr355del.

URLs: ChIP Enrichment Analysis, http://amp.pharm.mmsm.edu/lib/chea.jsp; Motif Activity Response Analysis, http://test.swissregulon.unibas.ch/cgi-bin/mara; Gene Set Enrichment Analysis, http://www.broadinstitute.org/gsea/index.jsp

A-5 SUPPLEMENTARY INFORMATION 2

Supplementary Table 14. Primers for generation and sequencing of *GATA2* ZF2 domain. Underlined are the *Nde*I *and Bgl*II restriction endonuclease sites used for cloning.

Primer name	Purpose	Primer sequence (5' → 3')	Ta
GATA2-ZF2-F	PCR,cloning & colony PCR	GTACATATGCAGAACCGACCAC TC	58°C
GATA2-ZF2-V2-R	PCR & cloning	GTAAGATCTTCACTATTATTTC TTGCTCTTCTTG	
T7-R	- Colony PCR - Sequencing	GCTAGTTATTGCTCAGCGG	-PCR 58°C -Standard 60°C

APPENDIX B: Additional Information for Chapter 4

B-1 STATEMENT/DECLARATION OF CONTRIBUTION

Chan Eng Chong (Candidate):

1) Performed mutagenesis on *GATA2* (mutants) expression constructs, bacterial transformation, plasmid extraction, prepared samples for sequencing and data analyses.

2) Prepared all plasmids including FLAG-*GATA2* expression constructs and FLAG-*GATA2* lentivirus constructs, performed bacterial transformation, DNA extraction, prepared DNA samples for sequencing and data analyses.

3) Performed all tissue culture works (HEK293, HEK293T, HL-60 and Cos-7), transduction of *GATA2* into HL-60, protein expression and extraction, immunofluorescence staining, luciferase reporter assay and data analysis on all samples, performed the experiment design and data interpretation.

4) Protein blot analyses (Western blotting and WEMSA) were performed in collaboration with Young Kyung Lee. Young also helped in protein extraction and tissues cultures.

5) Performed all protein structural modellings.

6) Prepared *Gata2* retrovirus constructs.

7) Bone marrow extraction and clonogenic assay were performed in collaboration with Grant Engler.

8) Animal study and experiment design and data interpretation were performed by Anna Brown.

I hereby certify that the statement of contribution is accurate.

Signed Date 15/4/13

Young Kyung Lee (Research Assistant-Molecular Pathology Research Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 15/4/13

Jan Kazenwadel (Collaborator -Lymphatic Development Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 15/4/13

Natasha Harvey (Collaborator and Head of the Lymphatic Development Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 16.4.13

Grant Engler (Research Assistant -Acute Leukaemia Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 16/4/13

Anna Brown (Collaborator -Acute Leukaemia Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 16/4/13

NOTE:
This appendix is included on pages 176-177 of the print copy of the thesis held in the University of Adelaide Library.

APPENDIX C: Additional Information for Chapter 5

C-1 STATEMENT/DECLARATION OF CONTRIBUTION

Chan Eng Chong (Candidate):

1) Performed preliminary studies on transactivation of GATA2 on *Prox1*. Prepared all plasmids including mutagenesis of *Gata2* (mouse) expression constructs and *PEE* luciferase reporter constructs, performed plasmid extraction, prepared DNA samples for sequencing, performed all sequencing data analyses, tissue culture works, protein expressions and immunofluorescence staining.

2) Protein blot analyses (Western blotting and WEMSA) were performed in collaboration with Young Kyung Lee. Young also helped in tissue culture works and protein extractions.

3) Luciferase reporter assays were performed in collaboration with Jan Kazenwadel.

4) Cloning of *Prox1* -11.3kb and sub-cloning of *PEE* were performed by Jan Kazenwadel.

I hereby certify that the statement of contribution is accurate.

Signed Date 15/4/13

Young Kyung Lee (Research Assistant -Molecular Pathology Research Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 15/4/13

Jan Kazenwadel (Collaborator -Lymphatic Development Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 15/4/13

Natasha Harvey (Collaborator and Head of the Lymphatic Development Laboratory)

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the results and data in the thesis.

Signed Date 16.4.13

C-2 SUPPLEMENTARY INFORMATION 1

Supplementary Table 1. Primers for *Gata2* (mouse) mutagenesis. Nucleotide changes leading to missense mutations (bold underlined). Nucleotides deletion is indicated by **|**.

Primer Name	Primer Sequence (5' → 3')
mT354M-F	GCACCTGTTGTGCAAATTGTCAGA**T**GACAACCACCACCTT
mT354M-R	AAGGTGGTGGTTGTC**A**TCTGACAATTTGCACAACAGGTGC
mT355del-F	ATTGTCAGACG**|**ACCACCACCTTATGGCGCCGGAACGCCAACGGGG
mT355del-R	CCCCGTTGGCGTTCCGGCGCCATAAGGTGGTGGT**|**CGTCTGACAAT
mL359V-F	ATTGTCAGACGACAACCACCACC**G**TATGGCGCCGGAACGCCAACGGGGA CCCTGTGT
mL359V-R	ACACAGGGTCCCCGTTGGCGTTCCGGCGCCATA**C**GGTGGTGGTTGTCGT CTGACAAT
mR361L-F	CACCACCTTATGGC**T**CCGGAACGCCAACG
mR361L-R	CGTTGGCGTTCCGG**A**GCCATAAGGTGGTG
mR362Q-F	ATTGTCAGACGACAACCACCACCTTATGGCGCC**A**GAACGCCAACGGGGA CCCTGTGT
mR362Q-R	ACACAGGGTCCCCGTTGGCGTTC**T**GGCGCCATAAGGTGGTGGTTGTCGT CTGACAAT
mC373R-F	CTGTGTGCAACGCC**C**GTGGCCTCTACTAC
mC373R-R	GTAGTAGAGGCCAC**G**GGCGTTGCACACAC
mR398W-F	CCAGACCCGGAAT**T**GGAAGATGTCCAG
mR398W-R	CTGGACATCTTCC**A**ATTCCGGGTCTGG

Supplementary Table 2. Primers for *PEE* mutagenesis. GATA2 binding motif WGATAR (underlined) in *Prox1* enhancer/promoter element is mutated to from GA to CT (bold)

Primer Name	Primer Sequence (5' → 3')
mProx1_GA2617CT-F	AAGCAAACGCGAAGTTCAATATCA**CT**TAAACCCGGCATG AAG
mProx1_GA2617CT-R	CTTCATGCCGGGTT**TA**AGTGATATTGAACTTCGCGTTTG CTT
mProx1_GA3616CT-F	TTCTTATAATCTGCCATTTCTTCTGAA**CT**TAAACAGTAT CATTTTAGGCATTTGTG
mProx1_GA3616CT-R	CACAAATGCCTAAAATGATACTGTT**TA**AGTTCAGAAGAA ATGGCAGATTATAAGAA

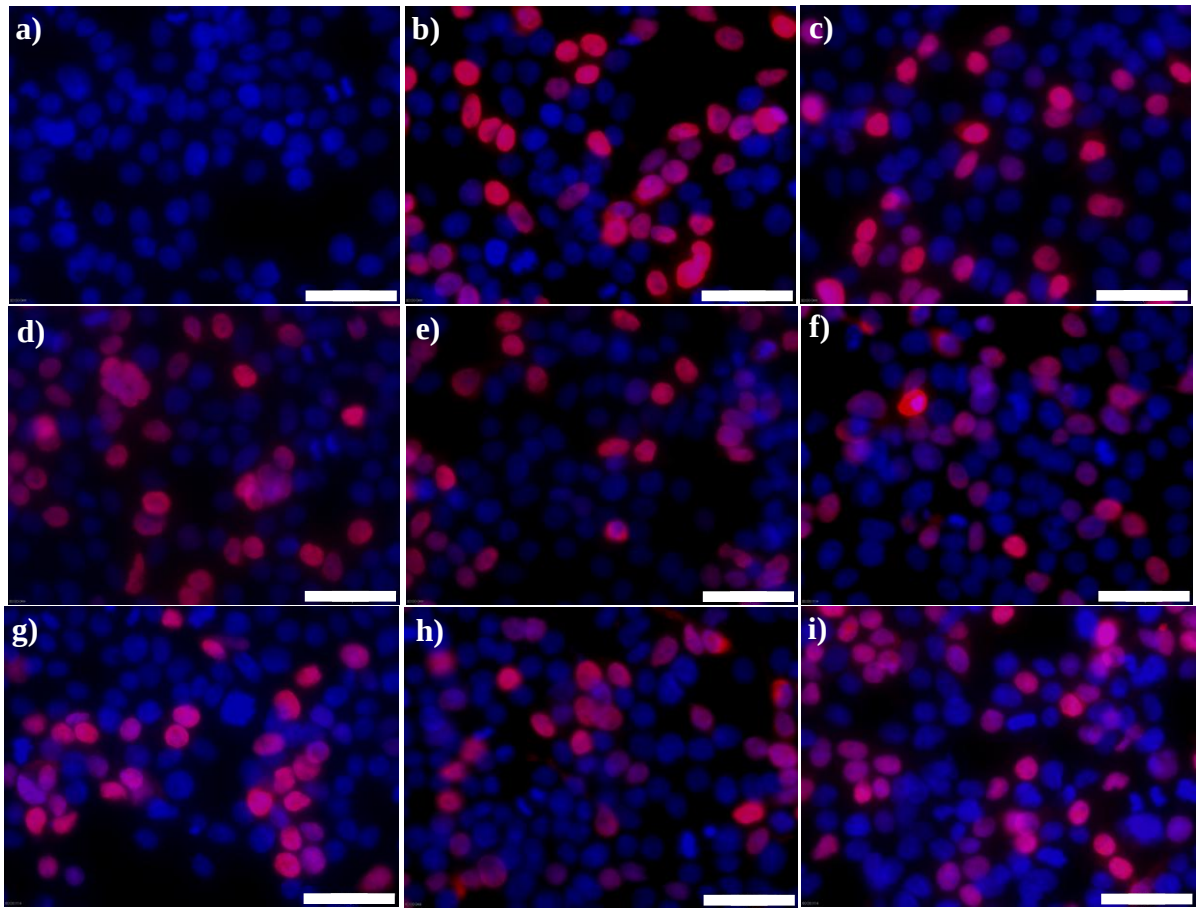
Supplementary Table 3. Sequencing primer for *PEE* luciferase reporter construct.

Primer Name	Primer Sequence (5' → 3')	Ta
2430-F	GCCTATTTGCATGGGTAGTCA	Standard 60°C
2783-R	GGAAACACCCTTGCCTACAT	
3423-F	CAGCAACAGAAAAGCCACAG	
3768-R	TCATGAACTCTGCTTGAAATGAA	

Supplementary Table 4. Primers for Western blotting-electromobility shift assay (WEMSA). GATA binding sites (underlined). Mutated binding site (underlined and bold).

Primer Name	Primer Sequence (5' → 3')
mProx1-32-F	TCAAATAATTTTATCAACCTGGACAG
mProx1-32-R	CTGTCCAGGTTGATAAAATTATTTGA
mProx1-121-F	TTTCTCTGAACTATCTGGATAATTAC
mProx1-121-R	GTAATTATCCAGATAGTTCAGAGAAA
mProx1-173-F	TCTGGGCTCCAGATAGATGTAGTAAA
mProx1-173-R	TTTACTACATCTATCTGGAGCCCAGA
mProx1-2593-F	GTTCAATATCAGATAAACCCGGCATG
mProx1-2593-R	CATGCCGGGTTTAAGTGATATTGAAC
mProx1-3592-F	TTTCTTCTGAAGATAAACAGTATCAT
mProx1-3592-R	ATGATACTGTTTAAGTTCAGAAGAAA
mProx1-2593-mut-F	GTTCAATATCA**CT**TAAACCCGGCATG
mProx1-2593-mut-R	CATGCCGGGTT**TA**AGTGATATTGAAC
mProx1-3592-mut-F	TTTCTTCTGAA**CT**TAAACAGTATCAT
mProx1-3592-mut-R	ATGATACTGTT**TA**AGTTCAGAAGAAA

C-3 SUPPLEMENTARY INFORMATION 2



Supplementary Figure 1. Expression of mouse GATA2 mutants in HEK293 cells.

Immunofluorescence staining of GATA2 using anti-GATA2 antibodies and Alexa 555-labelled secondary antibodies. The cells were stained for GATA2 (pink) and DAPI (blue). Scale bars, 50 μm, 400X magnification. a) Empty vector (pCMV6-Entry), b) GATA2 WT, c) T354M, d) T355del, e) L359V, f) R361L, g) R362Q, h) C373R and h) R398W.

Supplementary References

1. Brown, A.L., et al., *Genetic regulators of myelopoiesis and leukemic signaling identified by gene profiling and linear modeling.* Journal of leukocyte biology, 2006. **80**(2): p. 433-47.
2. Tsuzuki, S., et al., *Potentiation of GATA-2 activity through interactions with the promyelocytic leukemia protein (PML) and the t(15;17)-generated PML-retinoic acid receptor alpha oncoprotein.* Mol Cell Biol, 2000. **20**(17): p. 6276-86.
3. Diessenbacher, P., et al., *NF-kappaB inhibition reveals differential mechanisms of TNF versus TRAIL-induced apoptosis upstream or at the level of caspase-8 activation independent of cIAP2.* J Invest Dermatol, 2008. **128**(5): p. 1134-47.
4. Callus, B.A., et al., *Triggering of apoptosis by Puma is determined by the threshold set by prosurvival Bcl-2 family proteins.* Journal of molecular biology, 2008. **384**(2): p. 313-23.
5. Koldej, R., et al., *Optimisation of a multipartite human immunodeficiency virus based vector system; control of virus infectivity and large-scale production.* J Gene Med, 2005. **7**(11): p. 1390-9.
6. Kumar, R., et al., *CBFA2T3-ZNF652 corepressor complex regulates transcription of the E-box gene HEB.* J Biol Chem, 2008. **283**(27): p. 19026-38.
7. Bates, D.L., et al., *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, 2008. **381**(5): p. 1292-306.
8. Crossley, M., M. Merika, and S.H. Orkin, *Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains.* Mol Cell Biol, 1995. **15**(5): p. 2448-56.
9. Zhang, P., et al., *Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8705-10.
10. Freson, K., et al., *Molecular cloning and characterization of the GATA1 cofactor human FOG1 and assessment of its binding to GATA1 proteins carrying D218 substitutions.* Hum Genet, 2003. **112**(1): p. 42-9.
11. Tevosian, S.G., et al., *FOG-2: A novel GATA-family cofactor related to multitype zinc-finger proteins Friend of GATA-1 and U-shaped.* Proc Natl Acad Sci U S A, 1999. **96**(3): p. 950-5.
12. Lecuyer, E., et al., *The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1.* Blood, 2002. **100**(7): p. 2430-40.
13. Tsuzuki, S. and T. Enver, *Interactions of GATA-2 with the promyelocytic leukemia zinc finger (PLZF) protein, its homologue FAZF, and the t(11;17)-generated PLZF-retinoic acid receptor alpha oncoprotein.* Blood, 2002. **99**(9): p. 3404-10.
14. Tsuzuki, S., et al., *Cross talk between retinoic acid signaling and transcription factor GATA-2.* Mol Cell Biol, 2004. **24**(15): p. 6824-36.
15. Pimanda, J.E., et al., *Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development.* Proc Natl Acad Sci U S A, 2007. **104**(45): p. 17692-7.

16. Qiu, Z., et al., *GATA transcription factors regulate the expression of the human eosinophil-derived neurotoxin (RNase 2) gene.* J Biol Chem, 2009. **284**(19): p. 13099-109.
17. Nottingham, W.T., et al., *Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer.* Blood, 2007. **110**(13): p. 4188-97.
18. Soliera, A.R., et al., *Transcriptional repression of c-Myb and GATA-2 is involved in the biologic effects of C/EBPalpha in p210BCR/ABL-expressing cells.* Blood, 2008. **112**(5): p. 1942-50.
19. Huang, Z., et al., *GATA-2 reinforces megakaryocyte development in the absence of GATA-1.* Mol Cell Biol, 2009. **29**(18): p. 5168-80.
20. Kitajima, K., et al., *Redirecting differentiation of hematopoietic progenitors by a transcription factor, GATA-2.* Blood, 2006. **107**(5): p. 1857-63.
21. Gaines, P., et al., *GATA-1- and FOG-dependent activation of megakaryocytic alpha IIB gene expression.* J Biol Chem, 2000. **275**(44): p. 34114-21.
22. Tavor, S., et al., *Restoration of C/EBPalpha expression in a BCR-ABL+ cell line induces terminal granulocytic differentiation.* J Biol Chem, 2003. **278**(52): p. 52651-9.
23. Dorsam, S.T., et al., *The transcriptome of the leukemogenic homeoprotein HOXA9 in human hematopoietic cells.* Blood, 2004. **103**(5): p. 1676-84.
24. Shimahara, A., et al., *Acetylation of lysine 564 adjacent to the C-terminal binding protein-binding motif in EVI1 is crucial for transcriptional activation of GATA2.* J Biol Chem, 2010. **285**(22): p. 16967-77.
25. Robert-Moreno, A., et al., *Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1.* The EMBO journal, 2008. **27**(13): p. 1886-95.
26. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers.* Methods Mol Biol, 2000. **132**: p. 365-86.
27. Robinson, C.W., et al., *Dermal necrosis and chromosome Iq abnormality in a man with a familial myeloproliferative disorder.* Aust. N. Z. J. Med., 1983. **13**(2): p. 141-145.
28. Capon, F., et al., *Fine mapping of the PSORS4 psoriasis susceptibility region on chromosome 1q21.* J. Invest. Dermatol., 2001. **116**(5): p. 728-30.
29. Horwitz, M., et al., *A family inheriting different subtypes of acute myelogenous leukemia.* Am.J.Hematol., 1996. **52**(4): p. 295-304.
30. Bates, D.L., et al., *Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA.* J. Mol. Biol., 2008. **381**(5): p. 1292-1306.
31. Liang, M.P., et al., *WebFEATURE: An interactive web tool for identifying and visualizing functional sites on macromolecular structures.* Nucleic Acids Res., 2003. **31**(13): p. 3324-3327.
32. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-64.
33. Bengtsson, H., et al., *aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory*, 2008, Department of Statistics, University of California: Berkeley.
34. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.

35. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
36. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* J Royal Stat Soc., 1995. **57**(1): p. 289-300.