Identification and Characterisation of Genetic Lesions that Predispose to and Gene Expression Patterns that Contribute to Myeloid Malignancies

PARVATHY VENUGOPAL
Bachelor of Science (Zoology with Medical Microbiology)
Master of Science (Regenerative Medicine)

A thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences
Discipline of Genetics and Evolution

Faculty of Sciences
The University of Adelaide, South Australia

March 2016

Table of Contents

Abstract

Acute Myeloid Leukaemia (AML) is a heterogeneous disease caused by multiple genetic lesions. Our laboratory focuses on understanding the genetics of both inherited and acquired haematopoietic malignancies. In this thesis, I have investigated both inherited and acquired genetic changes that contribute to myeloid malignancies.

One of the key factors regulating haematopoiesis is GATA2, a zinc finger transcription factor. Germline mutations in *GATA2* have been associated with several clinical phenotypes such as myelodysplastic syndrome (MDS)/AML, immunodeficiency disorders (MonoMAC syndrome, DCML deficiency, congenital neutropenia, NK cell deficiency, aplastic anaemia) and Emberger syndrome. Moreover, several somatic mutations in *GATA2* have been reported in MDS/AML. Intriguingly, missense somatic and germline mutations reported to date are mutually exclusive, and several clinical phenotypes are associated with specific mutations. We generated a zinc finger 2 (ZF2) mutant allelic series representing a range of clinical phenotypes to investigate how each mutation effects transactivation, DNA binding, protein structure, protein partner interactions and *in vitro* differentiation. Specific GATA2 mutations perturb the interactions and functions in distinct ways that are beginning to explain differences in observed clinical phenotypes.

We performed gene expression analysis of 91 selected MDS/AML genes, including *GATA2*, on 166 well annotated primary AML samples, bone marrow mononuclear cells (BMMNC) and CD34 controls. Correlation analyses of *GATA2* expression levels with expression of other genes and other mutational and clinical data, was performed to help identify genetic aberrations that cooperate with abnormal levels of *GATA2* in AML.

Statistical correlations of expression levels of various other genes with outcome and mutation status were also identified.

One such correlation was reduced *GATA2* expression with oncogenic RAS mutations. A pilot study was carried out to evaluate and optimise an NRAS G12D-induced leukaemia model. All mice transplanted with mutant *NRAS* G12D rapidly developed haematopoietic disease post-transplantation whereas the control group did not. Based on these pilot studies, we have initiated transplantation experiments in a conditional *GATA2* knockout model to investigate the requirement of *GATA2* in NRAS G12D induced myeloid disease. Recipient mice continue to be monitored, but are yet to develop disease.

We also identified gene expression patterns of prognostic significance in AML and narrowed down a combination of three genes that are highly predictive of outcome. We devised a strategy integrating these genes into currently used risk stratification strategies and significantly improved risk stratification of AML patients at diagnosis.

Among syndromes that predispose to MDS/AML, is Diamond Blackfan Anaemia (DBA), a congenital disorder characterised by red blood cell deficiency. The underlying genetic cause of DBA in a child was identified using whole genome sequencing (WGS), targeted massively parallel sequencing (MPS) and high density SNP array. A complex scenario of germline and somatic aberrations were identified in two genetic loci that helped to explain the clinical features seen in the patient and the progression of this disease. These have led to the discovery of a mechanism by which spontaneous remissions occur in DBA patients.

Together, these studies have given us valuable insights into malignant myeloid disease biology and offer potential applications in improving therapeutic approaches in AML patients.

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 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,

Parvathy Venugopal

31 March 2016

Acknowledgements

I wish to thank my supervisors Hamish Scott, Chris Hahn and Manuela Klingler-Hoffmann for providing guidance and support throughout my PhD. Thank you for taking time to give me valuable feedback on my thesis in spite of being busy with deadlines and a million other things. The opportunity to work in the Department of Genetics and Molecular Pathology has been invaluable and I have learnt so much in my time here.

I would like to express my sincere gratitude to Hamish for giving me this opportunity and always encouraging me to step out of my comfort zone. I think that is some of the best advice I will ever get. I admire his ability to see the big picture and his brutal honesty which gets the best out of you. He is an incredible individual who is passionate about making a difference in people's lives through his work.

I'm extremely grateful to Chris for his invaluable assistance and encouragement over the last three years. Chris personifies patience, always making time and explaining things as many times and in as many ways as it would take for one to understand. I cannot emphasise enough on the integral role he has played in my PhD. Thank you Chris!

My sincere thanks to Manuela for helping me get started in the lab and giving me that first push I needed to ease into a work environment/culture that was completely new to me. Her role in helping me settle in my early days was immense and I am grateful for that.

I would like to thank all the Scott lab members- Milly, Pete (special thanks for all his help with looking after the mice), Anna, Chan-Eng, Young, Bergithe, Bradley, Jesse and Alicia for their friendship, company and support in all things scientific and non-scientific. Thank you Milly, Alicia and Jesse, for ensuring I didn't go hungry by buying me lunch or chocolate in both solid and liquid form.

Special thanks to Richard D'Andrea, Anna Brown, Mahmoud Bassal, Justine Marum, Jim Manavis, John Finnie, Andreas Schreiber, Joel Geoghegan, David Lawrence, Jinghua (Frank) Feng, Michele Grimbaldeston, Dave Yip and all our collaborators for their inputs and contributions to various aspects of my work. Thanks to the patients and their families for their willingness to participate in research without which my thesis wouldn't have been complete. Special thanks to Justine Marum and Ranjini Vaidyanathan for helping

me make sense of random forests and other complex statistical methods used in Chapter 4. All members of the Genetics and Molecular Pathology department, the ACRF genome facility, the animal facility and the Detmold facility deserve special mention for their help and their kind and encouraging words on some of my not-so-great days.

I am grateful to my parents for letting me chase my dreams. I would like to thank my father for being supportive of unconventional career (and partner) choices, my mother for being my super-cool pillar of support and my sister for backing me in all my decisions. I thank my family and friends for being proud of my achievements and I will be forever indebted to my grandfather for encouraging me to ask questions and introducing me to the fascinating world of science.

Most importantly, I would like to express my everlasting gratitude to my partner, Vishal, for his unconditional love and support over the last decade and for patiently putting up with me during these last few months. I fall short of words to express how valuable his contribution to this entire process has been. Thank you for keeping me sane!

List of Figures

List of Tables

Abbreviations

ADBA	Australian Diamond Blackfan Anaemia
AGM	Aorta-gonad mesonephros
ALT	Alanine transaminase
AML	Acute Myeloid Leukaemia
APS	Ammonium persulfate
BFU-E	Burst forming unit-erythroid
BMCMCs	Bone marrow-cultured mast cells
BMMNC	Bone marrow mononuclear cell
BMT	Bone marrow transplant
BSA	Bovine serum albumin
CD	Circular dichroism
CFU-G	Colony forming unit-granulocyte
CFU-GEMM	Colony forming unit-granulocyte, erythroid, macrophage, megakaryocyte
CFU-GM	Colony forming unit-granulocyte, monocyte
CFU-M	Colony forming unit-monocyte
CML-BC	Chronic myeloid leukemia in blast crisis
CMML	Chronic myelomonocytic leukaemia
CMV	Cytomegalovirus
CN	Cytogenetically normal
Co-IP	Co-immunoprecipitation
DBA	Diamond Blackfan Anaemia
DCML	Dendritic cell, monocyte, B and NK lymphoid
DEL	Deletion
DEPC	Diethylpyrocarbonate
DFS	Disease Free Survival
DNP	Dinitro-phenyl-albumin
DSS	Disease Specific Survival
EDTA	Ethylenediaminetetraacetic acid
EFS	Event Free Survival
ELN	European LeukaemiaNet
EMSA	Electromobility shift assay

EV	Empty vector
FAB	French-American-British
FACS	Fluorescence Activated Cell Sorting
FBS	Foetal Bovine Serum
G-CSF	Granulocyte-colony stimulating factor
GFP	Green Fluorescent Protein
GMP	Granulocyte Monocyte Progenitor
GOF	Gain-of-function
H&E	Haematoxylin and Eosin
HEK	Human Embryonic Kydney
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
Het	Heterozygous
HSC	Haematopoietic Stem Cell
HSPC	Hematopoietic stem and progenitor cells
ID	Immunodeficiency disorders
IFC	Integrated Fluidic Circuit
IL-6	Interleukin 6
ITC	Isothermal titration calorimetry
JMML	Juvenile myelomonocytic leukaemia
LB	Luria Broth
LBI	*LAPTM4B/BSPRY/IDH1*
LCL	Lymphoblastic cell line
LOF	Loss-of-function
LOH	Loss of heterozygosity
LSK	$Lin^{-}Sca1^{+}c\text{-}Kit^{+}$
MDS	Myelodysplastic Syndrome
MLL	Mixed-lineage leukemia
MNC	Mononuclear cell
MOI	Multiplicity of infection
MonoMAC	Monocytopenia with *Mycobacterium avium* complex
MPD	Myeloproliferative disease
MPS	Massively Parallel Sequencing
MRI	Magnetic resonance imaging
MSCV	Murine stem cell virus

Mut	Mutant
NGS	Next Generation Sequencing
NK	Natural Killer
NLS	Nuclear Localisation Signal
NMR	Nuclear magnetic resonance
NSCLC	Non-Small Cell Lung Cancer
OS	Overall Survival
PAGE	Polyacrylamide gel electrophoresis
PBMNC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PBS-T	PBS-Tween 20
p-NAG	p-Nitrophenyl-N-Acetyl-β-D-Glucosaminide
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RA	Refractory anaemia
RAEB	Refractory anaemia with excess blasts
RAEB-T	Refractory anaemia with excess blasts in transformation
RARS	Refractory anaemia with ring sideroblasts
rfsrc	Random forests for survival, regression and classification
RIPA	Radio-Immunoprecipitation Assay
RP	Ribosomal proteins
SACRB	South Australian Cancer Research Biobank
SCF	Stem cell factor
SDS	Shwachman Diamond Syndrome
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
STA	Specific target amplification
TALL	T cell lymphoblastic leukaemia
t-AML	Therapy related AML
TGE	Tris-glycine-EDTA
TNF	Tumour Necrosis Factor
TPO	Thrombopoietin
UPD	Uniparental disomy
VIMP	Variable importance
WB	Western blot

WEMSA	Western blotting-electromobility shift assay
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
WT	Wild type
ZF1	Zinc finger 1
ZF2	

Chapter 1. Review of Literature

1.1 Haematopoiesis

The haematopoietic system is the most well studied stem cell model that has laid the foundation for our understanding of stem cell biology. Haematopoiesis is the process through which haematopoietic stem cells (HSC) give rise to the wide variety of cells that form the haematopoietic system. These include red blood cells, platelets, granulocytes, macrophages, mast cells, natural killer (NK) cells and B- and T-lymphocytes (Figure 1.1). Being terminally differentiated, many of these cells have relatively short life spans. Hence, it is necessary to have a constant pool of HSCs that give rise to all these cell types during the lifetime of an organism. It is estimated that an adult human produces around 1.5 million blood cells every second (1).

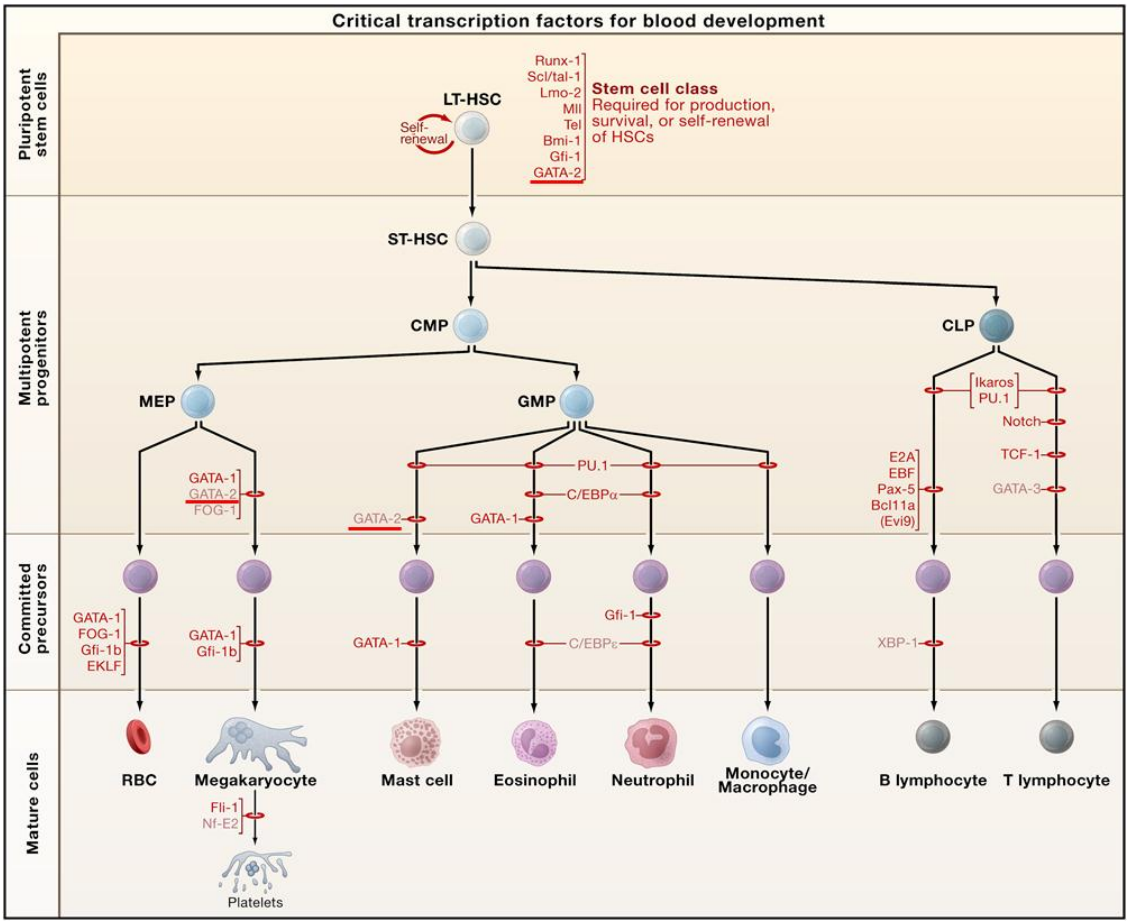


Figure 1.1 Haematopoiesis. Haematopoietic stem cells differentiate into all the mature cell types of the haematopoietic system in complex process orchestrated by a myriad of transcription factors and cytokines. (Adapted and modified from Orkin,2008) (2)

Haematopoeitic stem cells have the ability to undergo asymmetric cell division wherein they either give rise to identical daughter stem cells or differentiate into progenitor cells.

In haematopoiesis, HSCs reside on top of the developmental hierarchy and a progressive restriction is seen in differentiation potential as an HSC gives rise to a series of progenitor cells which eventually generate the diverse mature cell types that constitute the haematopoietic system (3). The organisation of this process into multiple levels accommodates for an increase in proliferative potential at particular stages in differentiation; and therefore results in an enormous number of differentiated cells being generated from each stem cell. For example, GMPs have a high proliferative index and can give rise to a large number of granulocytes. At the same time, HSCs are required to cycle infrequently and are therefore protected from the mutagenic effects of cell division.

1.1.1 Primitive Haematopoiesis

Haematopoiesis occurs at different anatomical locations during development, the earliest location being blood islands in the yolk sac where haematopoiesis takes place. This mainly involves production of red blood cells which is crucial for rapid growth at this stage. This is followed by pro-definitive (erythroid-myeloid progenitor), meso-definitive (erythroid-myeloid-lymphoid progenitor) and meta-definitive (neonatal repopulating HSC) haematopoiesis which have been shown to occur at several locations including the allantois, yolk sac, placenta and aorta-gonad mesonephros (AGM) (4). The AGM derived HSCs migrate to and colonise the foetal liver (FL) between E11.5-14.5 in mice embryos. HSCs migrate to the embryonic spleen at E12 and finally to embryonic bone marrow between E15.0 and 16.0 in preparation for birth. During early stages of development self-renewal is the predominant outcome of cell division. Cell cycling rate of HSCs reduces with age and haematopoiesis and maintenance of haematopoieisis becomes heavily reliant upon expansion of progenitor cells.

1.1.2 Regulators of Haematopoiesis

Several transcription factors are known to control early haematopoietic development as well as lineage commitment and differentiation including SCL/TAL1, GATA2, LMO2 and RUNX1. Several studies have shown that these factors are absolutely critical and absence can lead to severe defects which often result in embryonic lethality. Homozygous *Tal1* knockout embryos die at E9.5 due to complete absence of haematopoiesis in the yolk sac. *Gata1* and *Myb* levels are undetectable in these embryos (5). GATA2 has been shown to be essential for both primitive haematopoiesis in the yolk sac and definitive haematopoiesis in fetal liver, spleen and adult bone marrow (6, 7). LMO2 was found to be indispensable for primitive erythropoiesis and definitive haematopoiesis as shown in

Lmo2 null/wild type chimeric mice (8). Homozygous *Runx1* knockout mice completely lack fetal liver haematopoiesis (9). HOXB4 has been shown to be critical for differentiation to erythroid and granulocytic lineages *in vitro* (10). GATA1 and PU.1 are determinants of megakaryocytic–erythroid and myeloid lineage specification and appear to be activated early upon differentiation of HSCs (11, 12). Recent studies, which looked into the genome wide binding patterns of haematopoietic transcription factors revealed combinatorial interactions between at least seven different transcription factors (TAL1, LYL1, LMO2, GATA2, RUNX1, ERG, and FLI1) to regulate haematopoietic progenitor function(13). These findings suggest that lineage commitment and differentiation of HSCs is governed not by a single key regulator but by complex regulatory networks (Figure 1.2). Deregulation of this process often leads to disorders varying in severity from mild anaemia or thrombocytopaenia to overt leukaemia.

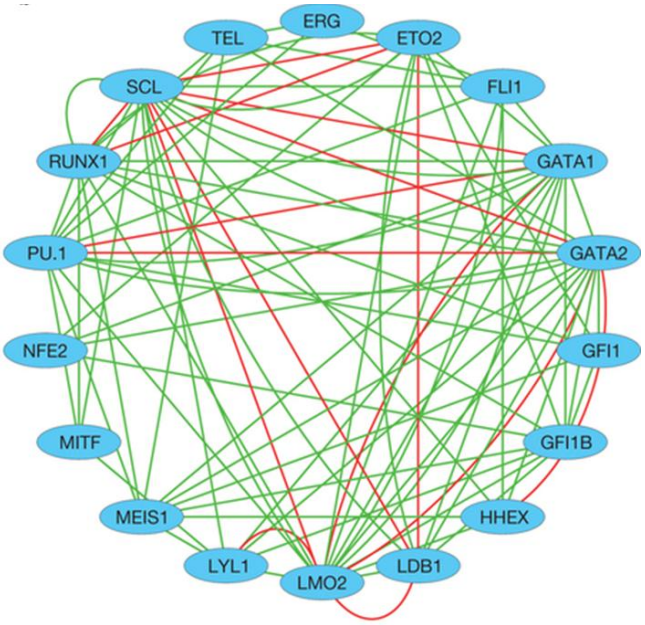


Figure 1.2 Complex regulatory network illustrating interactions between 18 core haematopoietic transcription factors. Green lines show known functional relationships and red lines show protein-protein interactions (14).

1.2 Myeloid Malignancies

Myeloid malignancies are disorders caused by genetic and epigenetic defects leading to aberrant differentiation and/or proliferation of haematopoietic stem cells or progenitor populations. This results in accumulation of abnormal cells within the bone marrow and peripheral blood. The genetic landscapes of myeloid malignancies have been studied extensively and many recurrent somatically mutated genes which initiate and drive disease have been identified.

1.2.1 Acute Myeloid Leukaemia

Acute myeloid leukemia is a malignant disorder characterised by clonal expansion of any one of the several non-lymphoid hematopoietic progenitor cells. It results from increased proliferation, survival and/or disruption of differentiation of haematopoietic precursors. As a result, there is an accumulation leukemic blasts in the marrow and peripheral blood. Patients with AML normally present with symptoms related to complications of pancytopaenia (*e.g.* anaemia, neutropaenia, and thrombocytopaenia), including weakness and fatigue, infections of variable severity, and/or haemorrhagic findings such as gingival bleeding (15).

A high proportion of children and adults (approx. 55%) with *de novo* AML and patients with therapy-related AML (t-AML) also carry chromosomal abnormalities. Cytogenetic findings constitute an independent prognostic factor. Major cytogenetic studies indicate that the prognosis of patients with inv(16)/t(16;16), t(8;21), and t(15;17) is relatively favorable as opposed to patients with inv(3)(q21q26) or t(3;3)(q21;q26) and monosomy 7 (16, 17). Aberrant expression of *EVI1* located at 3q26 has been reported to be a consequence of t(3;3)(q21;q26) and inv(3)(q21;q26) (18). However since the report of *GATA2* (located on 3q21.3) as a predisposition gene, these structural chanes have been studied more closely (19). Now it is believed that inv(3)(q21;q26) brings a *GATA2* distal haematopoietic enhancer into close proximity of *EVI1* driving its overexpression in haematopoietic stem and progenitor cells while also resulting in functional haploinsufficiency of *GATA2*, thereby contributing to leukaemogenesis (20, 21).

1.2.1.1 FAB Classification Scheme for AML

In 1976, French, British and American haematologists came together to propose the French-American-British (FAB) system to establish a uniform system for classification and nomenclature of acute leukaemias. The objective of the proposal was to define morphological criteria to distinguish between lymphoblastic and myeloid leukaemias at the time of diagnosis. As per the FAB system (which was revised in the 1980s), AML can be divided into 8 groups (M0-M7) listed in Table 1.1. The FAB system was used worldwide following that meeting.

Table 2.1 French-American-British Classification Scheme for AML

FAB subtype	Description
M0	Undifferentiated acute myeloblastic leukaemia
M1	Acute myeloblastic leukaemia with minimal maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute promyelocytic leukaemia (t (15:17))

M4 M4eos	Acute myelomonocytic leukaemia Acute myelomonocytic leukaemia with eosinophilia
M5a M5b	Acute monoblastic leukaemia Acute monocytic leukaemia
M6	Acute erythroid leukaemia
M7	Acute megakaryocytic leukaemia

1.2.1.2 World Health Organisation (WHO) Classification Scheme for AML

The WHO published a new classification scheme in 2001 to incorporate genetic, morphologic, immunophenotypic and clinical information into diagnostic criteria for AML (See Table 1.2). WHO classification has retained some aspects of FAB classification in the AML NOS (not otherwise specified) category. Though the WHO classification is more relevant to treatment and prognosis, the FAB classification is still widely used. A revision of the WHO classification is under way and the section on AML with recurrent genetic abnormalities is expected to change to accommodate recent findings in this area. A section recognising familial predisposition to myeloid neoplasms is also being developed (22).

Table 1.2 WHO classification scheme for MDS and AML (Adapted from Vardiman, 2010) (23)

Myelodysplastic/myeloproliferative neoplasms

- Chronic myelomonocytic leukaemia
- Atypical chronic myeloid leukaemia, BCR-ABL1 negative
- Juvenile myelomonocytic leukaemia
- Myelodysplastic/myeloproliferative neoplasm, unclassifiable
- Refractory anaemia with ring sideroblasts associated with marked thrombocytosis

Myelodysplastic syndromes

- Refractory cytopaenia with unlineage dysplasia
    - Refractory anaemia
    - Refractory neutropaenia
    - Refractory thrombocytopaenia
- Refractory anaemia with ring sideroblasts
- Refractory cytopaenia with multilineage dysplasia
- Refractory anaemia with excess blasts
- Myelodysplastic syndrome associated with isolated del(5q)

- Myelodysplastic syndrome, unclassifiable
- Childhood myelodysplastic syndrome
    - Refractory cytopaenia of childhood

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

AML with myelodysplasia-related changes

Therapy-related myeloid neoplasms

Acute myeloid leukaemia, NOS

- AML with minimal differentiation
- AML without maturation
- AML with maturation
- Acute myelomonocytic leukaemia
- Acute monoblastic and monocytic leukaemia
- Acute erythroid leukaemia
- Acute megakaryoblastic leukaemia
- Acute basophilic leukaemia
- Acute panmyelosis with myelofibrosis

Myeloid sarcoma

Myeloid proliferations related to Down syndrome

- Transient abnormal myelopoiesis
- Myeloid leukaemia associated with Down syndrome

Blastic plasmacytoid dendritic cell neoplasm

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 leukemia, B/myeloid, NOS
- Mixed phenotype acute leukaemia, T/myeloid, NOS
- Natural killer (NK) cell lymphoblastic leukaemia/lymphoma

1.2.2 Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS) are a group of disorders characterised by ineffective production of different myeloid lineages which often manifest as peripheral cytopaenias which coexist with a typically hypercellular and rarely hypocellular bone marrow. There exists considerable heterogeneity within MDS both clinically and biologically. Cytogenetics, morphologic features, peripheral blood parameters and molecular genetics together help define prognostically relevant groups. Though MDS is associated with increased risk of AML, a large proportion of morbidity and mortality associated with MDS can be attributed to the severity of the cytopaenias seen in the patients (24).

Table 1.3 FAB Classification for MDS (Adapted from Vardiman, 2012) (25)

Category	Dysplasia	% BM blasts
Refractory anemia (RA)	Erythroid	<5
Refractory anemia with ring sideroblasts (RARS)	Erythroid	<5
Refractory anemia with excess blasts (RAEB)	2 or more lineages	5–20
Refractory anemia with excess blasts in transformation (RAEB-T)	Usually 2 or more lineages	21–30
Chronic myelomonocytic leukemia (CMML)	Variable $\geq 1 \times 10^9$/L monocytes	<20

The FAB classification also provided criteria to identify and classify MDS on the basis of morphological features of blast cells (Table 1.3) (26). The WHO classification scheme (Table 1.2) integrated biologic, genetic and clinical information into durable framework of FAB classification for MDS. The major changes were lowering of blast percentages from 30% to 20% (for a diagnosis of MDS) which resulted in the elimination of the refractory anaemia with excess blasts in transformation (RAEB-T) category and

separation of chronic myelomonocytic leukaemia (CMML) into a new myelodysplastic/myeloproliferative neoplasms category (27).

1.3. Inherited Predisposition to Haematopoietic Malignancies

Inherited predisposition to haematopoietic malignancies within families has been known to exist for many years (28). It has largely been linked to syndromes whose effects are not limited to the haematopoietic system. A large number of inherited disorders like Emberger syndrome, Down syndrome, Fanconi anaemia, Bloom syndrome and Neurofibromatosis type 1 are known to predispose individuals to AML (29-31). AML is also associated with several acquired conditions including aplastic anaemia (32), MDS, acquired amegakaryocytic thrombocytopaenia (33, 34) and paroxysmal nocturnal haemoglobinuria. Non-syndromic familial inheritance of AML, though relatively rare, has been reported in families with mutations in *CEBPA*, *RUNX1* and *GATA2*. In light of recent findings, syndromic and non-syndromic AML are rapidly merging into a single group.

1.3.1. Familial Leukaemia

Familial leukaemia refers to high risk families that aren't linked to complex medical syndromes. Most predisposition genes described are often transcription factors known to play important roles in haematopoiesis. They often have variable penetrance and typically have an autosomal dominant mode of inheritance. Several of these genes are commonly mutated in sporadic AML as well.

RUNX1, which is often mutated, deleted, amplified or disrupted by gene fusions or chromosomal aberrations in sporadic AML, is also responsible for familial platelet disorder predisposed to AML (FPD/AML). This autosomal dominant disease is characterised by variable levels of platelet dysfunction which manifests as bleeding tendency in early childhood; and a lifelong risk of developing MDS/AML (35).

Germline mutations in *CEBPA* have been noted to segregate with familial leukaemia, often with normal karyotypes and a predominance of FAB subtypes M1 and M2, in an autosomal dominant pattern. Both sporadic and familial CEBPA mutations are seen to confer a good prognosis. Also, biallelic mutations are a common occurrence in CEBPA associated AML.

Germline mutations in *GATA2* have been identified in families with a predisposition to MDS/AML (19). Though it was first discovered as an MDS/AML predisposition gene, subsequent reports identified germline *GATA2* mutations/deletions in Emberger syndrome

and MonoMAC/DCML deficiency syndrome (36-38). Hence, GATA2 deficiency is no longer considered a pure familial leukaemia disorder.

Since then several different genes have been identified in familial haematological malignancies including *TERT/TERC, ANKRD26, SRP72, ACD, PAX5, ETV6* and *DDX41* (reviewed in (39)).

1.3.2. Syndromic Predisposition to Haematopoietic Malignancies

Several inherited syndromes are accompanied by a predisposition to haematopoietic malignancies among other non-haematopoietic manifestations. Causative genes for such syndromes usually have a wide range of developmental functions which is not limited to the haematopoietic system. Nevertheless, these syndromes provide us with very useful models to study the underlying mechanisms of leukaemogenesis and understand better the role of the causative genes in normal haematopoiesis.

1.3.2.1 Inherited Bone Marrow Failure Syndromes

Inherited bone marrow failure syndromes are a heterogeneous group of disorders characterised by bone marrow failure involving one or multiple lineages and frequently associated with physical anomalies and an increased risk of developing MDS/AML. Bone marrow failure syndromes have been linked to defects in ribosome biogenesis, telomere maintenance and DNA repair pathways. Germline mutations in DNA repair genes causes defective repair mechanisms and leads to diseases like Fanconi Anaemia (*FANC)* which have an underlying predisposition to different types of leukaemia. Mutations in components of ribosome biogenesis cause Shwachman Diamond Syndrome (*SBDS*) and Diamond Blackfan Anaemia (*RPS19, RPS24, RPS17, RPL35A, RPL5, RPL11, RPS7, RPL36, RPS15, RPS27A, RPS26*) (40, 41).

1.3.2.3 Tumour Suppressor Gene Syndromes

Mutations in tumour suppressor genes usually follow an autosomal dominant inheritance. Li Fraumeni syndrome is caused by germline mutations in *TP53* and is characterised by a predisposition to many different kinds of malignancies including osteosarcoma, breast cancer, brain tumours and leukaemia among others (42). Neurofibromatosis is caused by germline heterozygous alterations to *NF1* and is accompanied by an increased risk of juvenile myelomonocytic leukaemia (JMML) or AML as well as neuroectodermal tumours, gastrointestinal stromal tumours, rhabdomyosarcoma among others.

1.4 GATA2 is an MDS/AML Predisposition Gene

1.4.1 Structure and Properties of GATA2

The GATA family of transcription factors (GATA1-6) is crucial for differentiation and maintenance of a broad range of cells and tissues. DNA binding domains of all the GATA transcription factors are highly conserved and recognise the WGATAR motif via their zinc fingers. The human *GATA2* gene is located on chromosome 3 and has 6 exons. Transcription of *GATA2* is initiated from the distal first exon (IS) in haematopoietic and neuronal cells and from the proximal first exon (IG) in all other tissues which express *GATA2* (43). Nevertheless translation begins in the common exon 2, thereby generating identical proteins from both promoters. The GATA2 protein consists of two zinc finger domains (ZF1 and ZF2), N-terminal and C-terminal transactivation domains, a nuclear localisation signal and a negative regulatory domain (44). GATA2 has been shown to interact with FOG1 through ZF1 whereas interaction with PU.1 occurs via ZF2 (45, 46). Both ZF1 and ZF2 have been shown to be essential for interactions with HDAC, RARα, PLZF and PIASy (47-50).

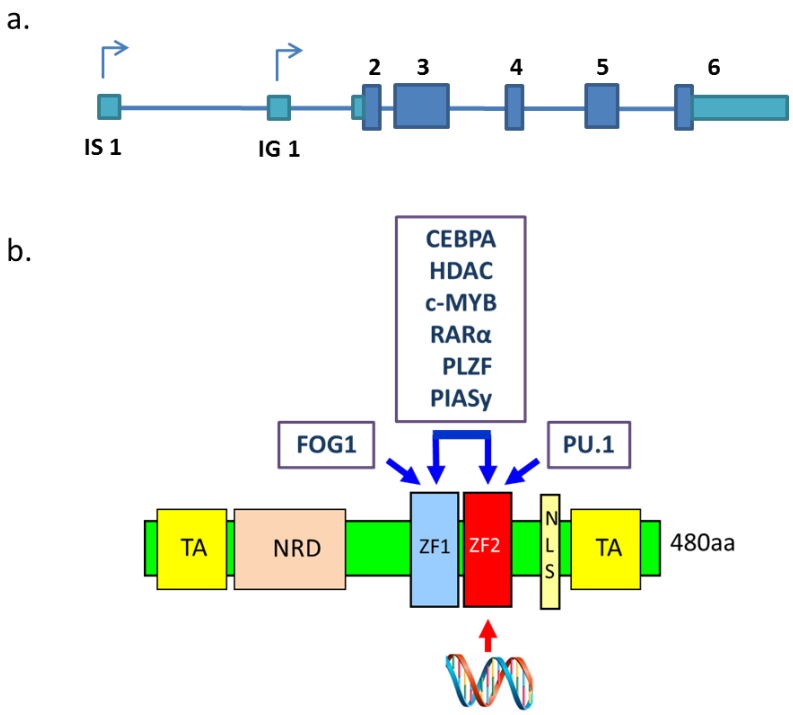


Figure 1.3 Schematic representation of the GATA2 gene and protein and known interacting proteins. (a) The GATA2 locus has 6 exons (dark blue). 5' and 3' untranslated regions have been shown in light blue. Transcription start sites are indicated by arrows at the IS and IG first exons. The protein has two transactivation domains (TA), two zinc finger domains (ZF1 and ZF2), a negative regulatory domain (NRD) and a nuclear localisation signal (NLS). Known protein interactions through the zinc finger domains are also shown.

1.4.2 GATA2 in Haematopoiesis

GATA2 is expressed within adult and developing haematopoietic stem cells, myeloid progenitors, erythroid precursors and mast cells. It is required for survival and homeostasis of immature haematopoietic progenitors and mast cell formation, but not for terminal erythroid and myeloid differentiation (6, 51). Gata2 knockout mouse embryos do not survive beyond E11 due to severe anaemia (6). Mice heterozygous for *Gata2* have significantly less stem cells. They have a larger proportion of quiescent cells and perform poorly in competitive reconstitution assays (51). They also show significantly reduced granulocyte macrophage progenitor (GMP) function while other committed haematopoietic progenitor cells such as common myeloid progenitors (CMPs) and lymphoid progenitors are unaffected (52). With respect to adult haematopoiesis, enforced expression of *Gata2* in haematopoietic progenitors reduces their number by 40-fold and hampers colony forming efficiency. Upon transplantation, primitive Sca-1^{+} Lin^{-} cells overexpressing *Gata2* showed markedly reduced capacity for amplification and differentiation (53). An optimum level of *Gata2* expression seems to be critical in maintenance and differentiation of HSCs, but the molecular mechanisms through which it acts are still unclear.

1.4.2.1 GATA2 in Erythroid Differentiation

Though *GATA2* is highly expressed in erythroid precursors, it is dispensable for terminal erythroid differentiation (54). *GATA2* and *GATA1* often show reciprocal expression patterns during erythroid development. *GATA1* binds to the region upstream of the *GATA2* promoter thereby repressing its expression during terminal erythroid differentiation. In the absence of *GATA1* expression, GATA2 can bind at this locus upregulating its own expression. The switch from GATA2 to GATA1 is an important determinant of erythroid commitment and differentiation and is commonly referred to as the GATA switch (44, 55).

1.4.2.2 GATA2 in Mast Cell Differentiation

Mast cells are derived from haematopoietic stem cells and their maturation takes place in tissues. Combinatorial expression of several transcription factors like GATA1, GATA2 and PU.1 are known to be involved in mast cell differentiation. Heterozygous knockouts of *Gata1* which survive to adulthood have lower number of mast cells than wildtype mice. Tsai et al demonstrated that *Gata2* is required for the formation of mast cells (54). There is high *Gata2* expression in early mast cells possibly indicative of its role in early mast cell gene expression. Presence of GATA consensus sequences in the promoter region of carboxypeptidase A, the α and β chain of the human IgE receptor (FcεRα and

β), IL-4, and IL-13 suggest that its role is not restricted to mast cell differentiation alone. There is evidence that repression of GATA2 activity causes impaired cell survival, IgE induced degranulation and anaphylactic response. Analogous to the GATA switch in erythroid differentiation, a similar switch has been described in mast cell differentiation where *FOG-1* expression in mast cell precursors blocks their maturation into mast cells by acting as a negative regulator of GATA2 and enabling GATA1 to take its place (2).

1.4.3 Aberrant GATA2 in MDS/AML

High GATA2 expression has been reported in AML and has been correlated with poor prognosis in adult and paediatric AML (56-58). Several studies have shown that *GATA2* expression levels were significantly high in many (MDS) and AML patients (59, 60). Hahn *et al* identified a highly specific heritable missense mutation (p.Thr354Met) in *GATA2* co-segregating with early onset MDS/AML in three families. In all three families, there are no individuals with AML or MDS who did not also carry T354M (19). Additionally, they identified T355del in a family where both members carrying the mutation developed MDS. A recent report identified germline *GATA2* mutations to be the causative lesion in 15% (13/85) of advanced and 7% (28/426) of primary paediatric cases of MDS in their study (61).

1.4.4 Other Manifestations of GATA2 Mutations/Deletions

Mutations in *GATA2* (L359V, Δ341-346) have also been reported in blast crisis in Chronic Myeloid Leukemia (CML BC). The L359V mutant exhibited enhanced transactivation and inhibition of PU.1, a major regulator of haematopoiesis (62). The majority of patients with CML BC have cells that bear close resemblance to blasts seen in AML. This may point to similarities in the underlying mechanisms which lead to both conditions.

Hsu *et al* discovered mutations in GATA2 associated with MonoMAC syndrome which is characterised by monocytopaenia with predisposition to mycobacterial and other specific infections. These included recurring T354M and R398W mutations, both in the ZF2. Concurrently, patients with the very closely related disease, dendritic cell, monocyte, B and NK lymphoid (DCML) deficiency, also harboured GATA2 mutations (63, 64). GATA2 mutations were identified in patients with congenital neutropaenia which often evolves into MonoMAC syndrome and MDS/AML with poor prognosis (65). A recent study has identified mutations in highly conserved intronic regions of GATA2 in MonoMAC patients (66). Bigley *et al* proposed that these disorders are the result of a

stem cell defect which specifically affects mononuclear cell production. The involvement of GATA2 mutations adds value to this possibility as GATA2 is known to be crucial for HSC maintenance and granulocyte-macrophage progenitor function (52, 64).

Aplastic anaemia is a stem cell disorder characterised by hypocellularity of bone marrow, reduced haematopoiesis and peripheral pancytopaenia (67). Low expression of GATA2 in HSCs and mesenchymal stem cells of patients with aplastic anaemia has been reported (68, 69). Recently, heterozygous mutations have been identified in the regulatory regions of *GATA2* in patients (70). These reports suggest that decreased expression of *GATA2* causes defects in the stem cell compartment which may lead to aplastic anaemia.

Ostergaard *et al* reported eight *GATA2* mutations (6 indels causing frameshifts and premature termination and 2 missense mutations in the ZF2 - R361L and C373R) in patients with Emberger syndrome, which is characterised by lymphoedema in addition to predisposition to MDS and AML (36). C373R has been predicted to affect the structure of zinc finger, as it replaces one of the crucial zinc ion-coordinating cysteines with a positively charged arginine that would repel the positively charged zinc ion. R361L has been predicted to affect DNA binding using structural modeling (37). Kazenwadel *et al*. reported that some of the patients diagnosed with MDS/AML or MonoMAC with *GATA2* deletions or frameshift mutations, also had primary lymphoedema. On investigation of GATA2 expression during development, they found selective localisation in lymphatic vessels and valves, but not blood vessels suggesting involvement of GATA2 in the development of the lymphatic vasculature. This reveals a previously unknown role for GATA2 in lymphatic development.

Most of the mutations described in patients with lymphoedema are nonsense mutations, frameshift mutations or large gene deletions resulting in complete loss of function of one allele. As discussed earlier, even the missense mutations in these patients are most likely loss-of-function as indicated by C373R, where the cysteine in ZF2 is replaced by arginine most likely destroying its function. Whereas, all the mutations identified in patients with MDS/AML and/or immunodeficiencies are missense mutations which cluster around very specific residues mainly in ZF2 and are indicative of gain of function. Mutations in GATA2, whose expression levels are known to be critical in regulation of haematopoietic stem cells, presumably upset haematopoietic homeostasis.

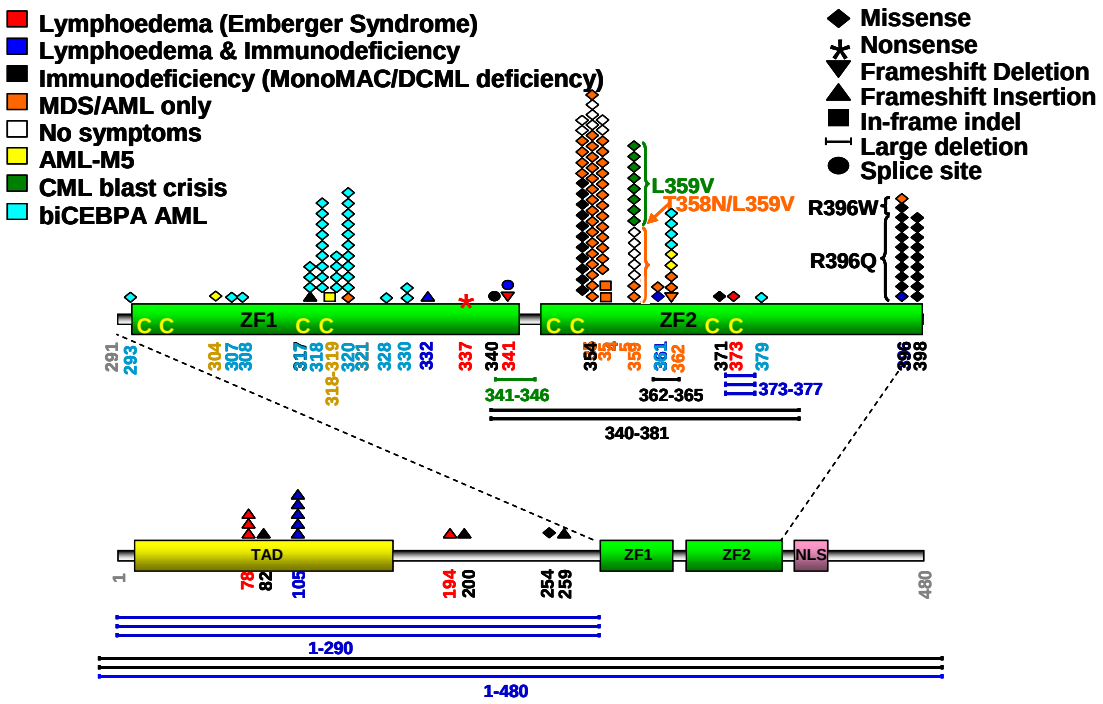


Figure 1.4 Germline missense mutations in *GATA2* cluster in ZF2 whereas somatic mutations cluster in ZF1. The above figure is a representation of different phenotypes and location of the associated mutated residue within the protein.

A striking common theme to all these different disorders is that one or more of the myeloid lineages (and sometimes lymphoid) are deficient or defective to varying degrees. This implicates a progenitor cell in the normal hierarchical classification as the root cause underlying the abnormality: the haematopoietic stem cell. Interestingly, there seems to be a common predisposition to MDS/AML associated with all these disease conditions. Despite common overall mechanisms leading to MDS and AML, the differences in phenotype observed between the patient cohorts and syndromes are likely to be a reflection of variations of activity conferred by each mutation – variations in localisation, stability and degradation leading to alterations in interaction with binding partners or DNA binding capacity or both. In Chapter 3, we shed more light on the molecular mechanisms that determine mutation-specific disease outcomes.

1.4.5 Somatic *GATA2* mutations in AML

Somatic *GATA2* mutations have been identified in 18-41% of AMLs with biallelic *CEBPA* mutations and 16% in monoallelic *CEBPA* mutated AML as well as in AML-M5 but only in 1-2% AMLs in total (71, 72). Interestingly, the majority of somatic *GATA2* mutations are missense mutations located within ZF1 with the exception of R362Q mutation (reported in AML-M5), and there haven't been any reports of missense germline mutations in ZF1 to date (Figure 1.4). In contrast germline *GATA2* mutations are highly

recurrent and cluster in specific residues in ZF2 which is highly suggestive of gain-of-function.

A summary of all *GATA2* mutations published till date has been included in Appendix III.

1.5 Gene Expression and Mutations in MDS and AML

The hallmarks of cancer as described by Hanahan and Weinberg are as follows: self-sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis, deregulation of cellular metabolism, evasion from the immune system, genomic instability and inflammatory microenvironment (73, 74). This suggested that there are crucial pathways whose function must be altered appropriately in order for cancer to develop. A mutation in any member of one of these pathways confers a cell with more than one of these characteristics thereby reducing the number of additional mutations/hits required for oncogenesis. There has been a practice of broadly separating genes mutated in haematopoietic malignancies into several functional categories (Figure 1.5)

Transcription factors – Mutations in haematopoietic transcription factors leading to impaired differentiation. For example, mutations in *CEBPA*, which is important for granulopoiesis, have been identified in 7-11% of AML patients (75, 76). *RUNX1* mutations have been reported in minimally differentiated AML and germline mutations leading to haploinsufficiency in *RUNX1* have been shown to predispose to AML in families (77). Also included are *GATA2*, *ETV6* and *PAX5*.

Tumour suppressor genes – Tumour suppressor genes can be altered by deletion, methylation or point mutations often resulting in blunted expression. For example, alterations in *TP53* occur frequently in t-AML with majority showing loss of the wild-type allele (78). p53 is considered the 'guardian of the genome' as it conserves genomic stability by halting cell cycle progression in response to DNA damage.

DNA Methylation Genes – DNA methylation genes like *IDH1/2*, *DNMT3A* and *TET2* are frequently mutated in MDS, leukaemia and lymphoma. AMLs with these mutations often show altered methylation patterns in comparison to non-mutant AMLs and/or healthy controls. *DNMT3A* mutant AMLs showed an overall reduction in CpG methylation compared to non-mutated AMLs (79). Loss of *TET2* has been shown to promote hypermethylation of enhancers associated with tumour-suppressor genes.

Chromatin modifiers – Mutations in chromatin modifiers like *ASXL1*, *EZH2* and *SUZ12* are thought to cause imbalance in temporal and lineage-specific gene regulation in

haematopoiesis. Large protein complexes such as the polycomb repressive complex (PRC) are crucial for normal regulation of this process and hence haematopoietic differentiation. *EZH2* and *SUZ12* are members of PRC2 whereas *ASXL1* is involved in recruiting PRC complexes and mediating interaction with DNA (80).

Spliceosome complex – Mutations in genes of the spliceosome complex have recently been identified in several haematopoietic malignancies including MDS, AML and CLL. The mutated components are largely involved in the initial steps of splicing. For example, *SF3B1*, which is frequently mutated in a particular subtype of MDS, encodes a core component of the U2 small nuclear ribonucleoprotein. SRSF2, which is involved in exon recognition and facilitates the recruitment of U2AF1 at the 3' splice site, is commonly mutated in CMML. We recently reviewed splice factor mutations in haematopoietic malignancies (see Appendix I) (81).

Signaling molecules – Mutations in signaling molecules usually provide cells with a proliferative and/or survival advantage. Mutations in *FLT3*, *NRAS*, *KRAS*, *NF1* and *KIT*, which are members of the RTK-RAS pathway, are frequently seen in AML (82). Somatic mutations in *FLT3* are found in almost 40% of AML patients which includes *FLT3* internal tandem duplication (FLT3-ITD) that accounts for about 30% and is often associated with poor prognosis (83).

Cohesin Complex – Components of the cohesin complex like *STAG2*, *SMC1A*, *SMC3* and *RAD21* have been found to be mutated in ~6% of AMLs (84). Cohesins are known to play a role in chromosome segregation, double strand DNA repair and transcriptional regulation. Mutations in cohesion complex genes possibly impair one or more of these functions and contribute to chromosomal instability.

For cytogenetically normal (CN) AML, a network of haematopoietic regulators including RUNX1, CEBPA, SPI1 (PU.1), FLT3, and others are strongly implicated by somatic mutation and over or underexpression in the leukaemogenic process as are as are signaling molecules common to many tumour types (TP53, NRAS, KRAS, HRAS, PTEN). More recently, epigenetic changes and mutations in the spliceosome machinery have been implicated in the oncogenic mechanism, adding another layer of complexity to understanding the underlying basis (85).

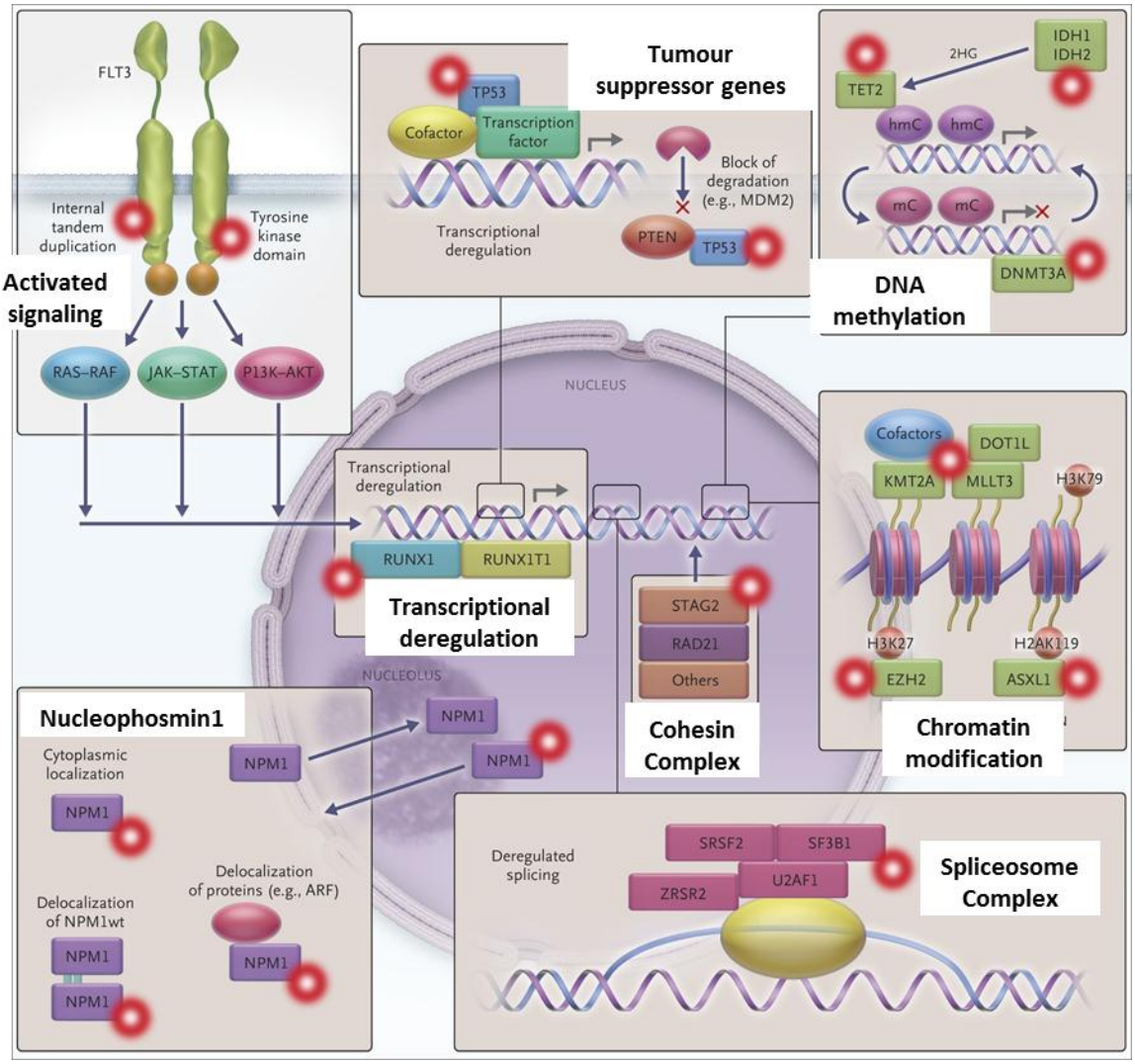


Figure 1.5 Recurrently mutated genes in AML can be separated into several functional categories. Recurrently mutated genes often confer a cell with one or more hallmarks of cancer. Adapted from Dohner *et al.* 2015 (22).

With the advent of next generation sequencing (NGS) technologies, the mutation profiles in MDS have been netter characterised and several recurrent mutations have been detected in MDS. The common mutations in MDS can be categorised into four groups – epigenetic modifiers (*ASXL1, TET2, DNMT3A, EZH2, IDH1, IDH2)*, RNA splicing genes (*SF3B1, SRSF2, U2AF1, ZRSR2*), signaling molecules (*NRAS, KRAS, JAK2*) and transcription factors (*RUNX1, TP53, ETV6*) (86).

1.6 Ribosomopathies

Amongst the syndromes that predispose to haematological malignancies are ribosomopathies. Ribosomopathies are defined as ‘any disease associated with a mutation in a ribosomal protein or biogenesis factor impairing ribosome biogenesis in which a defect in ribosome biogenesis or function can be clearly linked to disease causality’(87).

Ribosomes are ribonucleoprotein complexes that catalyse the essential cellular function of translation. Eukaryotic ribosomes are comprised of two subunits – the 40S subunit and 60S subunit - which together form the active 80S ribosome. Ribosome biogenesis is a complex process involving the assembly of over 70 ribosomal proteins. The process has high energy requirements and is reliant upon coordination of RNA polymerases. Contrary to expectations, the effects of mutations/deletions in ribosomal genes are not widespread but are restricted to specific tissue types. Ribosomopathies with a predisposition to MDS or AML have been outlined in Table 1.4.

1.6.1 Ribosome Dysfunction and Bone Marrow Failure

Due to high turnover of cells in the haematopoietic system, presumably haematopoietic stem and progenitor cells have a large requirement for ribosomes to cater to their metabolic needs. Hence, impaired ribosome biogenesis may lead to hypoproliferation in these cells giving rise to cytopaenias which are a common feature of marrow failure disorders. Also, improper assembly of ribosomes may lead to accumulation of free ribosomal proteins within the cells acting as a stress signal. Free ribosomal proteins have been described to bind and inhibit MDM2 that normally causes degradation of TP53. Inhibition of MDM2 can result in cell cycle arrest and thereby hypoproliferation, via TP53 stabilisation (41). It is still unclear how the switch from a hypoproliferative bone marrow failure to a hyperproliferative malignant phase takes place. One possible explanation for this switch is that the hypoproliferative state exerts selective pressure on cells which have a proliferative advantage which may be acquired either through revertant mutations or driver mutations. We discuss revertant mutations as a potential mechanism for auto-correction of a bone marrow failure phenotype in a DBA patient in Chapter 6.

1.6.1.1 Diamond Blackfan Anaemia

Diamond Blackfan Anaemia (DBA) is a bone marrow failure disorder characterised by red cell aplasia often accompanied by craniofacial abnormalities, thumb or upper limb abnormalities, cardiac defects, urogenital malformations and a high risk of MDS/AML. It is usually treated by steroid therapy and/or red cell transfusions. Patients often require iron chelation therapy to counter the effects of transfusion associated iron overload from improper recycling of iron from red cells. Heterozygous mutations or deletion of genes encoding ribosomal proteins have been identified to be causative of DBA. Both large and small subunit ribosomal proteins have been reported in patients including *RPL5, RPL11, RPL35A, RPS7, RPS10, RPS17, RPS19, RPS24, RPS26* among others (Figure 1.6) (88).

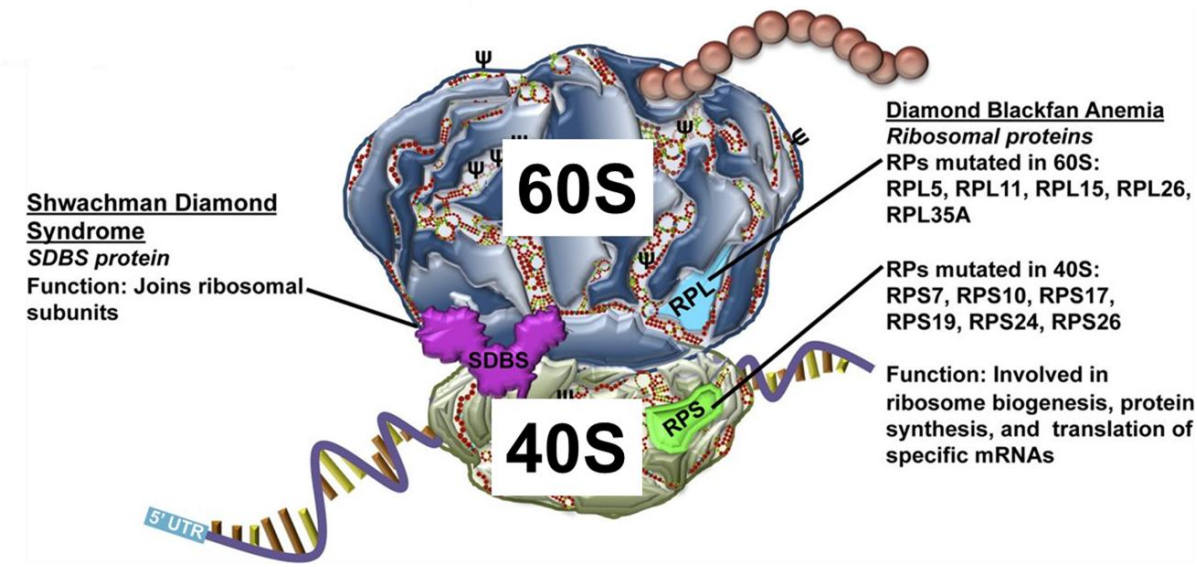


Figure 1.6 Key ribosomal components mutated in DBA and SDS. Adapted from Ruggero *et al.* 2014

1.6.1.2 Shwachman-Diamond Syndrome

Shwachman-Diamond Syndrome (SDS) is characterised by marrow failure and exocrine pancreatic dysfunction and often presents as neutropaenia, anaemia or thrombocytopaenia. SDS patients have a marked propensity to develop MDS/AML. SDS is caused by compound heterozygous mutations in *SBDS* with most mutations likely to be resulting from gene conversion from the adjacent highly homologous pseudogene. SBDS is required for the release of eIF6 from the 60S subunit which is an essential step in formation of a functional 80S ribosome (41).

Table 1.4 Ribosomopathies with a predisposition to MDS/AML

Disease	Gene	Clinical features	Cancer risk
Diamond Blackfan Anaemia (DBA)	*RPS19, RPS24, RPS17, RPL35A, RPL5, RPL11, RPS7, RPL36, RPS15, RPS27A, RPS26*	craniofacial abnormalities, thumb or upper limb abnormalities, cardiac defects, urogenital malformations, cleft palate	MDS/AML osteosarcoma
5q-Syndrome	Deletion of 5q, *RPS14*	macrocytic anemia, thrombocytosis, hypolobulated micromegakaryocytes	AML
Shwachman-Diamond syndrome (SBDS)	*SBDS*	Neutropaenia, infections, pancreatic insufficiency, short stature	MDS/AML
Dyskeratosis congenita	*DKC1*	Cytopaenia, skin hyperpigmentation, nail dystrophy, oral leukoplakia	AML, head and neck tumors

Adapted from Narla *et al.* 2010 (89)

1.7 Aim of the study:

Our understanding of causative lesions in sporadic and inherited predispositions to haematopoietic malignancies has improved tremendously with the advent of technologies like gene expression profiling and next generation sequencing. It is only in the last two decades that several familial predisposition genes have been discovered and we are yet to unravel how specificities of particular defects contribute to initiation, progression and maintenance of the oncogenic state.

In Chapter 3, we study the functional relevance of selected mutations in *GATA2* that are associated with specific myeloid malignancy phenotypes. We hypothesise that the differences observed in clinical presentation are a consequence of specific changes in structure and function of GATA2 protein conferred by each mutation.

In Chapter 4, we investigate gene expression of 91 genes selected for their importance in normal and malignant haematopoiesis which includes known haematopoietic transcription factors, genes of prognostic importance, downstream targets and interacting partners of GATA2 and epigenetic modifiers in a well annotated cohort of AMLs. We identify correlations between gene expression and mutation status as well as gene signatures that are of importance in determining prognosis.

In Chapter 5, we investigate recent findings that point to a link between oncogenic RAS and its dependence on *GATA2* expression. We use a mouse model of malignant haematopoietic disease to help to begin to understand this relationship. Preliminary results are discussed.

Finally, in Chapter 6, we achieve a molecular diagnosis for a particularly complex case of DBA, an inherited bone marrow failure disorder with an underlying predisposition to MDS/AML, utilising next generation sequencing and high density single nucleotide polymorphism arrays.

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.

Abcam - Sapphire Bioscience Pty Ltd, Waterloo, NSW, Australia.

Agilent - Integrated Sciences, Sydney, NSW, Australia.

Ambion® - Applied Biosystems, Scoresby, VIC, Australia.

Applied Biosystems - Life Technologies Australia Pty Ltd, Mulgrave, VIC, Australia.

BD Biosciences - BD Australia head office, North Ryde, NSW, Australia.

Biolegend - 9727 Pacific Heights Blvd, San Diego, CA 92121

Bio-Rad - Bio-Rad Laboratories (Pacific) Pty Ltd, Gladesville, NSW, Australia.

Bio-Tek Instruments, Inc - Millennium Science, Mulgrave, VIC, Australia.

Cell Signaling Technology - Genesearch Pty Ltd, Arundel, QLD, Australia.

Clonetech - Scientifix Pty Ltd, Clayton, VIC, Australia.

eBioscience – Headquarters, 10255 Science Center Drive, San Diego, CA 92121 USA

GE Healthcare - GE Healthcare Bio-Sciences Pty Ltd, Rydalmere, NSW, Australia.

GeneWorks - GeneWorks, Hindmarsh, SA, Australia.

Invitrogen - Invitrogen Australia Pty Ltd, Mulgrave, VIC, Australia.

Kodak - Kodak, Sydney, NSW, Australia.

Millipore - Merck Pty Ltd, Kilsyth, VIC, Australia.

New England Biolabs (NEB) - Genesearch Pty Ltd, QLD, Australia.

Olympus - Olympus Australia Pty Ltd, Richmond, SA, Australia.

Peprotech - Abacus ALS Australia, Brisbane, QLD, Australia.

Pierce - Thermo Fisher Scientific, Scoresby, VIC, Australia.

Promega - Promega Corporation, Alexandria, NSW, Australia.

QIAGEN - QAGEN Pty Ltd, Doncaster, VIC, Australia.

Roche - Roche Diagnostics Australia Pty Ltd, Castle Hill, NSW, Australia.

Santa Cruz - Quantum Scientific, Murarrie, QLD, Australia.

Sigma - Sigma-Aldrich Pty Ltd, Castle Hill, NSW, Australia.

STEMCELL™ Technologies - STEMCELL Technologies Australia Pty Ltd, Tullamarine, VIC, Australia.

Stratagene - Integrated Sciences Pty Ltd, Chatswood, NSW, Australia.

TAKARA - Scientifix Pty Ltd, Clayton, VIC, Australia.

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 Mini 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 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

2.2.4 Tissue Culture

Human embryonic kidney cell lines, HEK293T & HEK293, and monkey kidney fibroblast cell line, 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 was used to generate wildtype and mutant GATA2 expressing clones. These stably transduced cell lines were maintained in RPMI media 1640 (Invitrogen) with 10% FBS and 10µg/ml gentamicin.

2.2.5 Lipofectamine-Mediated Transfection

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

2.2.6 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).

2.3 Chapter 3 Methods

Details methods of **Chapter 3** were discussed in a manuscript that has been prepared for submission. Supplementary methods can be found in Appendix IV

2.3.1 Animal Handling and LSK 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 Committee (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, tibias, pelvises and humeri 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 red blood lysis buffer (QIAGEN). Total BM cells were then centrifuged (400xg, 10min), washed and counted in a haemocytometer. Subsequently, magnetic lineage depletion was performed using biotinylated lineage antibodies (TER119, B220, Mac1, CD3e, CD5 and Gr-1) and Dynabeads Biotin binder (Life technologies). This was followed by fluorescence activated cell sorting for Lin- Sca1+ cKit+ (LSK) cells using

BD FACSAria cell sorter (BD Biosciences). LSK cells were resuspended in StemSpan SFEM (Stem Cell Technologies) supplemented with 1X PenStrep-Glutamine (Invitrogen) and 100ng/ml each of the TPO, SCF and G-CSF (Peprotech), plated in retronectin (30 µg/ml, TAKARA) coated, non-tissue culture treated 24-well plates and cultured for 48 hours. The protocol was modified from Faber *et al* (personal communication Steven Lane) (90).

2.3.2 Generation of Retrovirus

Infectious retrovirus (ecotropic) was made by co-transfecting 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).

2.3.3 Retrovirus Transduction and Colony Forming Unit Assay

Retrovirus supernatant (EV, WT Gata2 or mutant Gata2) was added to non-tissue culture treated 24-well plates containing LSK cells (MOI 1). Subsequently, the plates were spun at 32°C, 1800 RPM for 60 minutes. The cells were then placed in incubator for 4h before a second similar transduction. After 24 h, the plates were topped up with fresh full media and cultured for another 24h. Successfully transduced LSK cells were sorted on GFP positivity, using the BD FACSAria cell sorter (BD Biosciences). For colony forming unit assay, GFP positive LSK cells were plated in 500 cells/ml of Methocult medium (STEMCELL™ technology) in the presence of G-CSF. All assays were performed in triplicate. After 7 days of culture, colonies were enumerated and typed under the light microscope (Olympus).

2.3.4 Bone marrow transplantation experiment

Recipient C57BL/6 mice aged 8-12 weeks were subjected to lethal irradiation at 9 Gy and allowed to recover for at least 4 hours before a minimum of 50,000 GFP positive cells along with 200,000 freshly isolated helper marrow cells were transplanted into them via tail vein injections. An estimate of the number of cells successfully injected into each mouse was recorded. Following transplantation, mice were monitored twice daily and weighed daily until day 21 after which they were weighed twice a week. Clinical record sheets were maintained to record and monitor their progress. Mice were bled to measure engraftment levels in peripheral blood.

2.3.5 *In vitro* mast cell differentiation

BMCMCs were obtained by culturing bone marrow cells from femurs and tibiae of *Tie2-Cre x Gata2* $^{flox/+}$ mice (wherein the floxed *Gata2* allele is knocked out in the haematopoietic and endothelial lineage) and their wildtype littermates in DMEM (Life Technologies) supplemented with 10% foetal calf serum (FCS; Bovogen) and 20% WEHI-3 conditioned medium (containing 3–4 ng/mL IL-3) for 4–6 wk, at which time > 95% of the cells were identified as mast cells by May Grünwald-Giemsa staining and by flow cytometric analysis (c-Kit+, FcεRI+) as described previously (91). All mast cell experiments were performed under the guidance of Michele Grimbaldeston and Dave Yip (Mast Cell Laboratory, Centre for Cancer Biology, Adelaide, Australia).

2.3.6 Detection of IL-6 and TNF secretion

IL-6 and TNF secretion was measured using the BD OptiEIA^{TM} Mouse IL-6 Kit and BD OptiEIA^{TM} Mouse TNF Kit (BD Biosciences) as per manufacturer's instructions.

2.3.7 β-hexosaminidase release assay

Mast cell degranulation was measured by β-hexosaminidase release assay. Briefly, BMCMCs were pre-sensitised with IgE (clone SPE-7) at 2 µg/ml and incubated overnight. Cells were washed in Tyrode's buffer and cell pellets were incubated with serial dilutions of DNP for an hour. Following the incubation, cells were centrifuged and the supernatant from each well was transferred to a different well. The remaining cell pellets were lysed in 0.5% Triton-X. p –Nitrophenyl-N-Acetyl-β-D-Glucosaminide (p-NAG) substrate was added to all wells followed by 0.2 M Glycine. The plates were read at 405 nm. The percentage of β-hexosaminidase released was calculated based on the amount recovered from the supernatant and the amount retained in the cell pellet.

2.4 Chapter 4 Methods

2.4.1 AML Patient Cohort

A panel of (n=166) AML patient samples was assembled using bone marrow mononuclear cell (MNC) preparations taken at diagnosis with consent via the South Australian Cancer Research Biobank (SACRB) with approval from relevant hospital ethics committees in accordance with the Declaration of Helsinki. Normal bone marrow samples were obtained using an approved collection process through the Royal Adelaide Hospital. Research ethics approval for this project was obtained through the Royal Adelaide Hospital human ethics committee

2.4.2 RNA Isolation, cDNA Preparation and Specific Target Amplification

Total RNA was isolated from 166 AML samples, 13 normal CD34 controls, 8 normal bone marrow controls and K562 cells using a routine phenol/chloroform method (Trizol, Invitrogen). cDNA was prepared from 250ng RNA using the Qiagen QuantiTect Reverse Transcription Kit as per manufacturer's protocol. Common mutations in AML were assessed. FLT3-ITD screening was performed by Bik To and Therapeutic Products Facility staff using PCR-based fragment analysis. KIT (D816V), DNMT3A (R882C/H), FLT3 (TKD: D835H/Y//V/E, I836DEL, I836INS), IDH1 (R132C/H/P), IDH2 (R140W/ L/G, R172W/G/K/M), JAK1 (T478S, V623A), JAK2 (V617F), KRAS (G12D/V/A, G13D/A), TET2, NPM1 and WT1 were assessed using a multiplexed matrix-assisted laser desorption/ionization time-of-flight genotyping approach (Sequenom MassARRAY Compact System, Sequenom, Inc., San Diego, CA, USA) performed by Andrew Wei and Nik Cummings (Department of Clinical Haematology, The Alfred Hospital and Monash University, Melbourne, Vic) in collaboration with Prof. Richard D'Andrea (Centre for Cancer Biology and University of South Australia, Adelaide, Australia).

Specific target amplification (STA) was used to increase the number of target cDNA. A 500 nM primer mixture (10x) were prepared by pooling 1 µL aliquots of all the primer pairs (100 µM) to be included in the STA reaction. 96 primer pairs were pooled, and the final volume adjusted to 200 µL. The pre-mix for the STA reaction was prepared by mixing 2.5 µL 2xTaqMan PreAmp Master Mix (Applied Biosystems), 0.5 µL 500nM (10x) pooled primer mix, and water to a total of 5 µL. 3.75 µL of the STA pre-mix were aliquoted for each sample (40 or 48), and added 2 µL°C DNA (~1 ng/µL). The samples were then amplified (95°C for 10 min, (95°C for 15 sec, 60°C for 4 min) x 14 cycles), 4°C). The STA samples were diluted 1:10 before use on the Dynamic Arrays.

2.4.3 96.96 Dynamic Arrays

The expression levels of the different transcripts in 166 AML samples, 13 normal CD34 controls and 8 normal bone marrow controls alongside K562 cells and water controls were measured utilizing two 96.96 Dynamic Arrays using EvaGreen on the Biomark HD System (Fluidigm). The Dynamic Arrays were run according to the manufacturer's protocol. Briefly, sample pre-mix were prepared using 2xSsoFast EvaGreen Supermix with Low ROX (Bio-Rad, PN 172-5211), and 20x DNA binding Dye Sample Loading Reagent (Fluidigm, PN 100-3738), both to a total concentration of 1x. 3.3 µL of this mix were added 2.7 µL of the diluted STA samples, a total of 96 reactions. The assay mix was

prepared by mixing 2.5 µL 2xAssay Loading Reagent, 2.2 µL 100 µM mix of the combined forward and reverse primers for each assay.

The Dynamic Array IFC was primed by loading control line fluid into each accumulator on the chip, before loading the chip into the IFC controller HX (96.96 and the Prime (136x) script). After the priming, 5 µL of each assay and 5 µL of each sample (48) were added to their respective inlets on the chip. Using the IFC controller software, Load Mix (136x) script was run.

The chips were run using the Data Collection Software using standard conditions: (70°C for 40 min, 60°C 30 s; only for the 96.96 Dynamic Array), 95°C for 60 s, (96°C for 5 s, 60°C for 20 s) x 30 cycles, 60°C for 3 s, hold at 60-95°C .

2.4.4 Data Output and Normalisation

The output file from the run named ChipRun.bml was loaded on Fluidigm Real Time PCR software. The sample and assay names were assigned. Baseline correction method was set at Linear (Derivative) and Ct threshold method was set at Auto (Global) prior to exporting raw data. The expression values for four housekeeping genes (B2M, HMBS, HPRT1 and RPLP0) for a random selection of samples was analysed using GeNorm which revealed the most and least stable housekeeping genes across samples. The data was normalised to the average of three most stable housekeeping genes (which eliminated B2M). Normalized data from both chip runs were combined for further analysis. Scatterplots were generated using R to analyse spread of expression of AML samples in comparison with normal controls. AMLs were also grouped based on mutation status to analyse differences in expression. Mann Whitney U test was performed to determine statistical significance.

2.4.5 Cohort Characteristics and Correlation Analyses

An oncoprint was generated in Microsoft Excel to better understand overlapping mutations in individual samples. Gene expression patterns correlating with gene expression patterns were identified. Statistical significance was calculated by Mann Whitney test. Pearson's correlation coefficients were calculated for combinations of genes across the samples using packages corrplot and cairo on R.

2.4.6 Kaplan Meier and Random forest analysis

Statistical significance was determined using log-rank test. Expression data for 85 genes for 122 AML samples along with overall survival, disease free survival, disease specific survival and event free survival data were analysed using R package – rfsrc (random

forests for survival, regression and classification) to simulate 1000 random classification trees. Variable importance (VIMP) measures are calculated and the genes are ranked based on importance for each variable. Genes found in the top ten important genes across all four outcomes were then used to generate a classification/decision tree. To avoid overfitting, this classification tree was then pruned by choosing a model with least cross validated error as determined by the complexity parameter (cp). Kaplan Meier analysis was carried using R and Graphpad Prism (GraphPad Software, Inc.) and log-rank test was used to determine statistical significance.

2.5 Chapter 5 Methods

2.5.1 Animal Handling and LSK Extraction

All experiments were performed according to protocols approved by the University of Adelaide Animal Ethics Committee (M-2015-266) and SA Pathology Animal Ethics Committee (39/14a).

All mice used in the study were either wildtype C57BL/6 or transgenics derived from this strain. C57Bl/6 Mx1-cre mice were obtained from Matthew McCormick (Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia) and C57Bl/6 Gata2 (flox) mice were obtained from Marja Salminen (University of Helsinki, Finland). C57Bl/6 Mx1-cre/Gata2(flox) strain was genotyped by PCR using the following primers: Cre Forward (CTG ACC GTA CAC CAA AAT TTG CCT G) and Cre Reverse (GAT AAT CGC GAA CAT CTT CAG GTT C) with a product size 210 bp; and Gata2(Flox) Forward (CTT TCC ACC CTC CTT GGA TT) and Gata2(Flox) Reverse (TTT TTC CCC AAA GTC ACC TG) with product sizes 505 bp for floxed and 471 bp for wildtype (92). 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, tibias, pelvises and humeri of 8- to 10-week-old female of appropriate genotypes were aseptically harvested and crushed using mortar and pestle in IMDM media (Invitrogen) supplemented with 15% FBS (Sigma). Red blood cells were depleted using red blood lysis buffer (QIAGEN). Total BM cells were then centrifuged (400xg, 10min), washed and counted in a haemocytometer. Subsequently, magnetic lineage depletion was performed using biotinylated lineage antibodies (TER119, B220, Mac1, CD3e, CD5 and Gr-1) and Dynabeads Biotin binder (Life technologies). Lineage depleted cells were resuspended in StemSpan SFEM (Stem Cell Technologies) supplemented with 1X PenStrep-Glutamine (Invitrogen) and IL-3 (30ng/ml), SCF (150ng/ml) and IL-6

(30ng/ml) (Peprotech), plated in retronectin (30 µg/ml, TAKARA) coated, non-tissue culture treated 6-well plates and cultured for 48 hours.

2.5.2 Generation of Retrovirus

The pMSCV-GFP-IRES-KRAS G12D constructs were obtained from Prof. Scott Lowe (MSKCC, New York, USA). Dr Chan-Eng Chong replaced the *KRAS* G12D with *NRAS* G12D. Infectious retrovirus (ecotropic) was made by cotransfecting HEK293T cells with pMSCV-GFP-IRES-NRAS G12D (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).

2.5.1 Retrovirus Transduction and Colony Forming Unit Assay

Retrovirus supernatant (EV, oncogenic *NRAS* G12D) was added to non-tissue culture treated 6-well plates containing lineage depleted cells (MOI 0.1). Subsequently, the plates were spun at 32°C, 1800 RPM for 60 minutes. The cells were then placed in incubator for 4h before a second similar transduction. After 24 h, the plates were topped up with fresh full media and cultured for another 24h. Successfully transduced lineage depleted bone marrow cells were sorted on GFP positivity, using the BD FACSAria II cell sorter (BD Biosciences) for colony forming unit assays. Cells were plated in 2500 cells/ml of Methocult medium (STEMCELL™ technology). All assays were performed in triplicate. After 7 days of culture, colonies were enumerated and typed under the light microscope (Olympus). Few representative colonies were picked, cytospun and stained with Giemsa to confirm colony type.

2.5.2 Transplantation:

Recipient C57BL/6 mice aged 10-12 weeks were subjected to lethal irradiation at 7.5 Gy and allowed to recover overnight before unsorted cells were transplanted into them via tail vein injections. An estimate of the number of cells successfully injected into each mouse was recorded.

For the pilot transplant experiment, mice were transplanted with 1,400,000 cells containing 3.5% or 7% NRAS G12D cells or 30% EV cells. Following transplantation, mice were monitored twice daily and weighed daily until day 21 after which they were weighed twice a week. Clinical record sheets were maintained to record and monitor their progress. Mice were bled at 3 weeks post transplantation and then at regular intervals to measure haematological parameters and GFP positivity. Mice were humanely killed if they lost more than 15% of their initial body weight or appear unwell. Peripheral blood,

bone marrow and spleen cells from these mice were analysed for GFP positivity by flow cytometry using Gallios (Beckman Coulter) or BD Fortessa (BD Biosciences). Sternum, lung, liver, spleen, thymus and lymph nodes were formalin fixed, paraffin embedded and H&E stained. Photographs of stained sections were captured on EVOS XL Cell Imaging System (Life Technologies).

For the GATA2/NRAS cooperation transplant experiment, lethally irradiated recipient mice were transplanted with 1,000,000 cells containing either <1% NRAS G12D cells or 20% EV cells. Mice were monitored twice daily and weighed daily until day 21 after which they were weighed twice a week. Mice were bled at 3 weeks post transplantation and then at regular intervals to measure haematological parameters and GFP positivity. Full blood counts are performed on the HemaVet alongside normal controls each time. Following the 3 week bleed, mice were given three intraperitoneal injections of Polyionsinic-Polycytidyle Acid (pIpC) at 12 µg/gm body weight, at 2 day intervals. DNA was isolated from peripheral blood at week 5 bleed to confirm excision of floxed GATA2 allele.

2.6 Chapter 6 Methods

Methods of **Chapter 6** were discussed briefly in a manuscript that has been prepared for submission. Supplementary methods can be found in Appendix X.

2.7 Media and Solutions

Values in parentheses are final concentrations. (Mostly relevant to methods used in Chapter 3).

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) - Stored at 4°C.

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

PBS-T (1000ml)

Phosphate buffered saline (PBS)	1000ml
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 1000 ml	

Tyrode's buffer (1000ml)

HEPES	10 ml	(1M)
NaCl	7.54 g	
KCl	0.37 g	
$CaCl_2$	0.206 g	
$MgCl_2$	0.203 g	
D-Glucose	1.008 g	
BSA	1 g	
Filtered water	Top up to 1000 ml	

Chapter 3. Clinically important driver mutations in GATA2 zinc finger 2 display functional diversity and tendency for specification of myeloid malignancy subtypes, immunodeficiency disorders and lymphedema

3.1 Introduction

Mutations in *GATA2* have been linked to MDS/AML, immunodeficiency and Emberger syndrome. To decipher disease mechanism, an allelic series of mutations in ZF2 of GATA2 representing the major disease phenotypes was generated. We performed functional assays to study the effect of each mutation on transactivation, DNA binding, protein structure, protein-protein interactions and *in vitro* differentiation.

3.2 Statement of Authorship

Statement of Authorship

Title of Paper	Clinically important driver mutations in GATA2 zinc finger 2 display functional diversity and tendency for specification of myeloid malignancy subtypes, immunodeficiency disorders and lymphedema.
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Parvathy Venugopal		
Contribution to the Paper	Optimization of protocols for ex vivo experiments. Organized, coordinated, performed and analysed data from cell isolation, tissue culture, retroviral transduction, clonogenicity and transplantation experiments. Luciferase assays. qRT PCR. Interpretation of data and contribution to writing manuscript.		
Overall percentage (%)	40%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Chan-Eng Chong		
Contribution to the Paper	Plasmid and retroviral preparation. Structural modelling. WEMSA. Luciferase assays. Analysis and interpretation of experimental and clinical data. Contributed to transplantation experiments and clonogenicity assays. Writing the manuscript.		
Signature		Date	24/02/2016

Name of Co-Author	Philippa H. Stokes		
Contribution to the Paper	ZF2 protein production and purification, EMSA, ITC. Contributed to writing the manuscript.		
Signature		Date	24/2/2016

Name of Co-Author	Young K. Lee		
Contribution to the Paper	Co-immunoprecipitation assays. qRT PCR. Contributed to cell isolation for transplantation experiments.		
Signature		Date	24/03/2016

Name of Co-Author	Peter J. Brautigan		
Contribution to the Paper	Plasmid preparation. Contributed to cell isolation and transplantation experiments.		
Signature		Date	25/02/2016

Name of Co-Author	David T.O. Yeung		
Contribution to the Paper	Clinical data anaylsis		
Signature		Date	4 MAR 2016

Name of Co-Author	Milena Babic		
Contribution to the Paper	Clonogenicity assays		
Signature		Date	25/02/16

Name of Co-Author	Grant A. Engler		
Contribution to the Paper	Clonogenicity assays		
Signature		Date	24/03/2016

Name of Co-Author	Steven W. Lane		
Contribution to the Paper	Expert advice on transplantation experiments		
Signature		Date	23 Feb 2016

Name of Co-Author	Manuela Klingler-Hoffmann		
Contribution to the Paper	Experimental planning, expert advice on protein experiments. Feedback on manuscript		
Signature		Date	29/02/16

Name of Co-Author	Jacqueline M. Matthews		
Contribution to the Paper	Experimental planning. Feedback on manuscript.		
Signature		Date	

Name of Co-Author	Richard J. D'Andrea		
Contribution to the Paper	Experimental planning. Feedback on manuscript.		
Signature		Date	26/2/16

Name of Co-Author	Anna L. Brown		
Contribution to the Paper	Experimental planning. Feedback on manuscript.		
Signature		Date	25/2/16

Name of Co-Author	Christopher N. Hahn		
Contribution to the Paper	Manuscript preparation. Experimental planning and data analysis. Luciferase assays. Feedback on manuscript.		
Signature		Date	25/02/2016

Name of Co-Author	Hamish S. Scott		
Contribution to the Paper	Manuscript preparation. Experimental planning and data analysis. Feedback on manuscript.		
Signature		Date	25/02/2016

3.3 Manuscript

Clinically important driver mutations in GATA2 zinc finger 2 display functional diversity and tendency for specification of myeloid malignancy subtypes, immunodeficiency disorders and lymphedema

Running title: FUNCTIONAL ANALYSIS OF GATA2 ZF2 MUTATIONS

Chan-Eng Chong,[1,2] Parvathy Venugopal,[1-3] Philippa H. Stokes,[4] Young K. Lee,[1,2] Peter J. Brautigan,[1,2] David T.O. Yeung,[1,2,5,6] Milena Babic,[1,2] Grant A. Engler,[5] Steven W. Lane,[7] Manuela Klingler-Hoffmann,[3] Jacqueline M. Matthews,[4] Richard J. D'Andrea,[2,5,8] Anna L. Brown,[1,2] Christopher N. Hahn,[1,2,6] and Hamish S. Scott[1-3,6,8,9#]

[1]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia;

[2]Centre for Cancer Biology, University of South Australia and SA Pathology, Adelaide, Australia;

[3]School of Molecular and Biomedical Science, University of Adelaide, Australia;

[4]School of Molecular Bioscience, University of Sydney, New South Wales, Australia;

[5]Division of Hematology, SA Pathology, Adelaide, Australia;

[6]School of Medicine, University of Adelaide, Adelaide, Australia;

[7]QIMR Berghofer Medical Research Institute, University of Queensland, Brisbane, Australia;

[8]School of Pharmacy and Medical Sciences, Division of Health Sciences, University of South Australia, Australia;

[9]Australian Cancer Research Foundation, Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, Australia

Author for correspondence

Professor Hamish S. Scott

Department of Genetics and Molecular Pathology

Centre for Cancer Biology

SA Pathology

PO Box 14, Rundle Mall,

Adelaide, 5000, South Australia

Australia

Phone: +61 8 8222 3651

Fax: +61 8 8222 3146

Email: hamish.scott@health.sa.gov.au

Keypoints:

- GATA2 zinc finger (ZF) 2 mutations direct specific subtypes of myeloid malignancies and/or other phenotypes.

Keywords: hematopoiesis; loss-of-function; mutation; GATA2; AML; MDS; lymphedema

Abstract: 210

Text: 4045

Figure count: 7

Table count: 0

Reference count: 66

Supplemental Figure count: 11

Supplemental Table count: 4

Abstract

Heterozygous germline *GATA2* mutations are associated with myeloid malignancies, immunodeficiency and lymphedema. A GATA2 mutant allelic series representing these major disease phenotypes caused by germline and somatic mutations in zinc finger (ZF) 2 was generated and functionally characterized. All mutants displayed reduced DNA binding which could be attributed to mutation of arginine residues critical for DNA binding or amino acids required for ZF2 domain structural integrity. A single exception was the sL359V mutation which showed WT DNA binding affinity but gain-of-function in some assays. Two GATA2 mutants bound PU.1 more strongly than wildtype (WT). Unlike GATA2 WT, all mutants were unable to suppress hematopoietic progenitor growth *in vitro*, allowing for colony expansion. GATA2 mutants also caused a shift from monocyte to granulocyte progenitor commitment. Analysis on patients with 3 recurrent germline (g) GATA2 mutations (gT354M, gR396Q and gR398W) revealed distinct clinical outcomes, with gT354M predicting the most aggressive disease course, followed by gR396Q and gR398W. Our studies are consistent with complete or partial GATA2 haploinsufficient loss-of-function predisposing to myeloid malignancy and/or immunodeficiency while permissiveness to lymphedema requires complete haploinsufficiency. Overall, domain- and amino acid-specific GATA2 mutations perturb the interactions and functions of this key myeloid transcription factor in distinct ways that are beginning to explain differences in observed clinical phenotypes.

Introduction

The association of GATA2 mutations in myeloid malignancies was initially reported in chronic myeloid leukemia – blast crisis (CML-BC) where ~10% of cases carried a mutation.[1] We subsequently reported heritable heterozygous *GATA2* germline (g) mutations, gT354M and gT355del, which transmit as an autosomal dominant inherited predisposition to developing myelodysplastic syndrome and/or acute myeloid leukemia (MDS/AML) with varying penetrance and disease onset.[2] These mutants possess compromised transcriptional activity due to reduced DNA binding and have dominant negative characteristics in some settings. Other than familial MDS/AML, patients harbouring germline GATA2 mutations also manifest an array of complex diseases with overlapping phenotypes including Emberger syndrome (primary lymphedema with MDS/AML),[3-5] immunodeficiency disorders (ID) such as MonoMAC syndrome,[6,7] DCML deficiency syndrome,[8,9] chronic neutropenia,[10] and NK cell deficiency,[11] and aplastic anemia,[12,13] all of which have an increased propensity to develop MDS/AML. Recently, a novel *in-cis* germline GATA2 mutation (gT358N;L359V) was reported in familial MDS/AML with thrombocytopenia.[14] Intriguingly, the change of one amino acid (T358N) completely nullifies a known somatic (s) gain-of-function (GOF) mutation in CML-BC (sL359V) resulting in overall loss-of-function (LOF).[15] This result was consistent with previous studies by us and others showing that germline LOF mutations lead to predisposition to MDS/AML and/or ID.[2,16,17] Patients with complete or partial gene deletions developed Emberger syndrome, further evidence that complete GATA2 haploinsufficiency is associated with predisposition to lymphedema.[3,5,16]

Somatic *GATA2* mutations are uncommon except in specific AML subtypes, *i.e.* AML-M5[18] and cytogenetically normal (CN) AML with mono- or biallelic *CEBPA* mutations.[19-

[22] Disease-associated somatic missense mutations encode for amino acid changes predominantly in the ZF1 of GATA2, with the exception of sR362Q in ZF2.[19-22] In contrast, no missense germline mutations occur within ZF1, rather all occur in the ZF2 domain and it's neigbouring C-terminal region.[2,6,7] Phenotypic variation and the unique mutational clustering separating germline and somatic *GATA2* mutations imply that they are functionally distinct from each other and represent at least two separate entities in which different cooperating mutations are favoured along the pathway to leukemia.[23] Furthermore, even within ZF2, missense mutations are broadly predisposing to MDS/AML, but appear to confer different clinical phenotypes. Thus, we hypothesized that the individual mutations share common aspects in MDS/AML development, but confer unique characteristics that lead to different disease subtypes. In this study, we generated an allelic series of ZF2 mutations in GATA2 representing major disease subtypes. Through *in vitro* functional characterization and genetic data analysis, we demonstrate there is a genotype-phenotype correlation whereby individual mutants confer specific and sometimes overlapping effects on GATA2 function. The commonalities and differences in GATA2 biological activities seen in these assays may explain the phenotypic diversity and pathogenic contribution of GATA2 ZF2 mutations.

Methods

Constructs

GATA2 expression constructs (human (*h*) pCMV6-XL6-GATA2 and mouse (*m*) pCMV6-Entry-GATA2) were purchased from OriGene. A *GATA2* mutant allelic series of expression constructs (human and mouse) were generated by site-directed mutagenesis (QuickChange, Stratagene) as described previously.[2] Tagged *h*GATA2 was generated by PCR using a 5' primer with FLAG tag sequence and subcloned into the pcDNA3 expression vector. A DNA fragment corresponding to the *hGATA2* cDNA ZF2 domain

(*i.e.*, residues 328-409) was subcloned into the pET11a bacterial expression vector (Novagen). For generation of retroviral expression constructs, *mGata2* WT and mutant cDNA coding regions were PCR amplified and subcloned into the pMSCV-IRES-GFP (Supplemental Table 1).

GATA2 ZF2 protein production, purification and characterisation

WT and mutant *h*GATA2 ZF2 proteins were expressed and purified as described,[24] with the exception that 20 mM Tris (pH 8.5) was maintained throughout the buffers, and only a single second cation exchange chromatography step, using a 50-600 mM NaCl step gradient was used. ZF2 proteins were dialysed into 20 mM sodium phosphate (pH 7.4), 50 mM NaCl, 1 mM DTT. Far-UV circular dichroism (CD) spectra (195-260 nm) were collected at protein concentrations of 5 μM using a JASCO J-815 CD spectropolarimeter at 25 °C. The resulting spectra were smoothed in Origin (Microcal) using five-point fast Fourier transform filtering. One-dimensional proton (1D-H^1) NMR spectra were collected at 25 °C in the same buffer on a Bruker AvanceIII 800 MHz spectrometer on samples (150-300 μM ZF2 protein) supplemented with 10% (v/v) D_2O, and 17 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid as an internal reference.

Isothermal titration calorimetry

Binding of WT and mutant *h*GATA2 ZF2 proteins (250-270 μM) to *hGM-CSF* oligonucleotide (20 μM) was performed as described[25] with minor modifications (Supplemental methods).

Cell isolation, retrovirus transduction and clonogenicity assay

Murine hematopoietic cells were isolated from femurs, tibias and hip bones of C57BL/6 mice and resuspended in Iscove's Modified Dulbecco's Medium (Invitrogen)

supplemented with 15% FBS. Erythrocytes were purged (RBC lysis buffer, Qiagen) before magnetic-bead depletion of lineage-positive (lin^+) cells (Dynabeads, Invitrogen). $\underline{L}in^-$ $\underline{S}ca1^+$ c-$\underline{K}it^+$ (LSK) cells were then selected using FACS sorting, (streptavidin-APC/Cy7, Scal-PE/Cy7 and c-Kit-APC, BioLegend®, 1:100 each; FACSAria™ II sorter, BD Biocsiences). LSK cells were cultured in retronectin (30 μg/ml, Takara) coated 24-well plates at a concentration of 0.5-1 × 10^5 cells/ml in StemSpan™ SFEM (StemCell Technologies) medium supplemented with SCF, TPO, G-CSF (100 ng/ml each, Peprotech) and Penicillin/Streptomycin/Glutamine (Sigma). Cells were expanded for two days before they were retrovirally spinfected twice (MOI=1 in each round) at 650 ×g, 32 °C for 60 min. The cells were cultured for 2 days before FACS-sorting. GFP^+ LSK cells were then seeded on a methylcellulose medium (MethoCult®, M3434, StemCell Technologies) containing the indicated cytokines. After 7 days of culture, colonies were enumerated and typed under a light microscope. For transplantation assays, 1 × 10^5 GFP^+ LSK cells (transduced with EV, WT, gT354M and gC373R) and 2 × 10^5 helper marrow cells were injected into the tail vein of lethally irradiated (1000 Rad) congenic mice. All animal experimental protocols were approved by the University of Adelaide and SA Pathology Animal Ethics Committees.

Patient information

Genetic and clinical information correlated with common germline GATA2 mutations (*i.e.*, gT354M, gR396Q and gR398W) was available from 96 patients;[2,3,6-11,16,17,26-32] gathered in familial studies from 27 confirmed kindreds and individuals with *de novo* mutations (Supplemental Table 2). All data was collected with informed consent with approval from Institutional Review Board/Human Research Ethics Committees.

Statistical analysis

All experiments were individually performed at least in triplicate. The non-parametric Mann-Whitney U test (GraphPad Prism 4) and Fisher's exact test (http://graphpad.com/quickcalcs/contingency1.cfm) were used to analyse statistical significance of observed differences. Survival and time to event analyses were performed using Kaplan-Meier, and Fine and Gray[33] methods, respectively, in R.

Results

GATA2 mutants display altered DNA binding

We generated a mutant allelic series for human GATA2 representing major disease phenotypes: MDS/AML (gT354M and gT355del),[2] CML-BC (sL359V),[1] Emberger syndrome (gR361L and gC373R),[5] acute monocytic leukemia (AML-M5),[18] biallellic *CEBPA* AML (sR362Q),[19,20,34,35] and MonoMac/DCML deficiency syndrome (immunodeficiency with MDS/AML, gR398W)[6-9] (Figure 1A). We also made corresponding mutant murine GATA2 clones. All mutant proteins displayed nuclear localisation similar to WT controls in HEK293 cells (Supplemental Figure 1).[2] We performed DNA binding assays using Western blotting-electromobility shift assay (WEMSA) with known GATA2-responsive elements from *hGM-CSF* and *hTCRD* enhancers (Figure 1B-D and Supplemental Figure 2). This analysis revealed a single GATA2-DNA complex for GATA2 WT. *m*GATA2 mutant proteins, as well as *h*GATA2 (Supplemental Figure 3), exhibited a very similar DNA binding pattern to both *hGM-CSF* and *hTCRD* probes. Consistent with the high degree of conservation at the amino acid level, we identified a range of relative DNA binding capabilities across the mutant allelic series, with all but sL359V displaying reduced apparent affinities. Notably, gR361L and gC373R showed little or no binding to DNA. Although reported previously,[2] we have included gT354M, gT355del and sL359V for comparison to other mutants.

We then performed a series of experiments to ascertain the mechanism by which each mutant impacted on DNA binding. This included isolating the ZF2 domain, and more accurately measuring DNA binding affinity, assessing structural integrity, and using *in silico* structural modelling. Thus, the ZF2 domain only of *h*GATA2 WT and mutants was expressed *in E. coli* and purified. DNA-binding affinities were assessed by electromobility shift assay (EMSA) and isothermal titration calorimetry (Figure 2 and Supplemental Figure 4). For this study, we also included the gR396Q, a mutant associated with congenital neutropenia,[10] MonoMAC,[6,7] MDS/AML[17,28,32] and Emberger syndrome.[30] The binding of ZF2 mutants to dsDNA corresponding to the *hGM-CSF* GATA2-responsive element revealed excellent concordance with that obtained from WEMSA, consistent with a previous report of the ZF2 domain or itsDNA-contact residues being major determinants of DNA binding affinity.[2] Based on far-UV CD spectropolarimetry and 1-dimensional ^{1}H-NMR spectroscopy (Supplemental Figure 5 and Kazenwadel *et al*. 2015[36]), the structure of the ZF2 domain of was not affected by mutation in most cases, with the exception that gC373R is poorly folded based on both methods, while gT354M and gT355del have WT levels of secondary structure (according to far-UV CD spectropolarimetry) but poor levels of tertiary structure (from NMR experiments). Zinc co-ordination is critical for structure in ZFs, so misfolding of gC373R is consistent with the loss of one of the 4 critical zinc co-ordinating cysteines (Figure 1A). A homology model of GATA2 ZF domains was generated based on a high resolution *h*GATA3-DNA structure[37] (*h*GATA3 is currently the closest homolog of *h*GATA2 for which a structure is known). The homology modelling indicates that mutation of positively charged arginine residues (R361 (black), R362 (red), R396 (yellow) and R398 (green)) likely disrupts critical interactions with DNA (Supplemental Figure 6). Our results are in line with previous modelling studies for gR396Q[38] and gR398W[9] which

used different GATA ZF templates. Collectively, these data indicate that alterations to the integrity of the ZF2 structure or mutation of key DNA-binding residues can severely impact on the ability of GATA2 to bind DNA.

GATA2 mutants display altered transactivation activity and protein-protein interaction

To determine the effect of mutations in GATA2 on the ability to transactivate transcription, we performed luciferase reporter assays on several known GATA2-responsive promoters (*LYL1* and *CSF1R*) and enhancers (*CD34*). All mutants displayed a reduced ability to transactivate transcription at these responsive elements, except for sL359V (Figure 3A, B, C and Supplemental Figure 7), which we and others have shown displays WT level or mild GOF activity.[1,2] For the *CD34* and *LYL1* constructs, the three mutants with least binding affinity for DNA (gT355del, gR361L, gC373R; Figure 1 and Figure 2) also showed the most marked reduction in transactivation, including complete abrogation of transactivation in some instances. Previously, we have shown that GATA2 and PU.1 (SPI1) synergistically upregulate *CSF1R* promoter activity.[2] Here, we show that although GATA2 mutants display reduced transactivation, they maintain an ability to synergise with PU.1 in *CSF1R* promoter activation (Figure 3C). This observation could be partially explained by altered protein-protein interaction. In co-immunoprecipitation experiments, all GATA2 mutants bound PU.1 (Figure 4 and Supplemental Figure 8). Strikingly, gT354M and gC373R consistently showed enhanced affinity for PU.1 compared to WT.

GATA2 mutants perturb differentiation of hematopoietic stem and progenitor cells

We investigated the effects of GATA2 mutations on hematopoiesis using colony forming assays after retroviral transduction of GATA2 mutants into murine LSK cells. GATA2

WT or mutant expression was confirmed using qRT-PCR (Supplemental Figure 9). All GATA2 mutants lost the potency of the WT to prevent clonal outgrowth (Figure 4A). Numeration of the types of colonies generated by the mutant allelic series revealed commonalities and striking differences (Figure 4B-C and Supplemental Figure 10). Common to all GATA2 (WT or mutants) expressing LSK cells was a significant reduction in the percentage of monocyte progenitors (CFU-M) (Figure 4B) with a concomitant increase in granulocyte (CFU-G; gT354M, gT355del, gR361L, sR362Q and gR398W) (Figure 4C) or erythroid progenitors (BFU-E; WT and sL359V) (Figure 4D). Noteworthily, three GATA2 mutants (gT354M, sR362Q and gR398W) with highest percentage CFU-G also show a reduction granulocyte-macrophage/granulocyte, erythrocyte, monocyte and megakaryocyte progenitor (CFU-GM/GEMM) (Figure 4E). Together, these results demonstrate that GATA2 ZF2 mutations impact on growth and differentiation programs in hematopoietic stem and progenitor cells (HSPC).

We also transplanted transduced LSK cells (empty vector (EV), GATA2 WT, gT354M or gC373R) into lethally irradiated recipient mice and monitored their steady-state hematopoiesis for up to 12 months (Supplemental Table 3 and Supplemental Figure 11). Consistent with the ability of GATA2 to block hematopoiesis in normal mice,[39] and to impair progenitor growth (above) we observed very low levels of short-term engraftment, and no long-term engraftment for GATA2 WT. In contrast, long-term engraftment was evident for vector-only or GATA2 mutants. Although we observed a range of short-term and long-term engraftment in a small number of recipients receiving GATA2 mutants, none of these mice developed hematological disorders. The numbers in this study are small, but suggest that the GATA2 mutants under study are not capable of transformation of HSPC in a background of normal levels of endogenous GATA2.

Recurrent *GATA2* mutations display unique clinical outcomes

To understand the clinical outcomes of the most common germline GATA2 mutation, we reviewed our in house clinical information and from the literature (n = 101, Supplemental Table 2). Five GATA2 patients for whom age of death were unknown were excluded from analysis. We then calculated survival and time to event data for 96 patients with the three most common mutations (gT354M, gR396Q and gR398W) (Figure 6A). In these analyses, the events of interest were the development of a hematological and/or an immunological disorder. Common to all mutations is the development of MDS with high penetrance. However, the age of MDS onset is significantly different, gR396Q patients being the earliest (median age 18) to develop an event, compared to gT354M and gR398W patients (median age of onset 32 and 40, respectively) (Figure 6B). Patients with the gR398W mutation usually present with an ID at around the same time as the initial diagnosis of the myeloid disorder (median age 39) as do about 40% of the gR396Q patients (Figure 6C). gT354M patients usually present with MDS/AML as the first sign of the clinical syndrome, which may mask or confound the diagnosis of immune defects (Figure 6C & D). AML specific survival is inferior in gT354M patients compared to others (Figure 6E). Interestingly, of 53 individuals with gT354M-driven hematopoietic disease, there were 30 AML (~57%) and no chronic myelomonocytic leukemia (CMML) cases while for 17 carriers of gR398W, there were only 3 cases of AML (~18%), but 6 of CMML (including a CMML-blast crisis) (~35%). No CMML was seen in the 21 gR396Q cases (Supplemental Table 4). This may imply a mutation-specific definition of disease subtype, even for mutations with similar *in vitro* DNA-binding, transactivation and clonogenic properties such as gT354M and gR398W. Notably, gT354M binds PU.1 stronger than WT, while gR398W binds with similar affinity as WT.

Discussion

Germline GATA2 mutations underlie the development of certain hematological, immunological and lymphatic disorders. Our investigation of the effect of recurrent and clinically important GATA2 amino acid changes show that changes to these critical residues significantly alter GATA2 function. Mutations in ZF2 do not overtly alter GATA2 nuclear localization, but they may perturb protein folding and/or DNA binding affinity. Homology modelling predicts that the 4 arginine residues (R361, R362, R396 and R398; Supplemental Figure 6) and L359[2] directly contact DNA. In contrast, T354, T355 and C373 are more likely to be important in maintaining structural integrity (Supplemental Figure 5). This is backed by protein folding and DNA binding affinity studies, noting that any protein whose folding is severely affected (*e.g.* gC373R) will also be unable to bind DNA. Hence, perturbation of DNA binding may disrupt GATA2 function in a way that promotes lymphedema, ID and/or myeloid malignancy.

Ali *et al*. propose 3 classes of GATA3 mutations based on DNA binding affinity.[40] We see similar DNA binding patterns for GATA2 mutants (Figure 7); enhanced, reduced and greatly reduced. Mutant sL359V belongs to Class I that displays WT equivalent or enhanced DNA binding affinity (Supplemental Figure 2). gT354M, sR362Q and gR398W are Class II partial LOF mutants with reduced DNA binding affinity. However, these mutants are still able to transactivate various GATA2-responsive promoters/enhancers although to a lesser extent. Class III consists of gT355del and Emberger syndrome mutants that do not bind (or bind very poorly) to DNA, and also include partial or whole gene deletions and premature termination mutants. Interestingly, mutations within each class lead to different clinical outcomes. The Class I mutant is a unique case of somatic mutation associated with driving CML into blast crisis. Class II mutants are associated with various myeloid malignancies and/or ID. In addition to these phenotypes, Class III

mutants also predispose to primary lymphedema which is consistent with our previous proposed model in which complete LOF of one GATA2 allele (entire GATA2 gene or promoter deletions) is required for predisposition to primary lymphedema.[3] This is further supported by our recent findings where heterozygous conditional knock-out of *Gata2* results in defective lymphatic vessel structure and vascular transport in adult mice.[36] In the stem-cell compartment, inactivation of one *Gata2* allele perturbs hematopoiesis in mice, and BM cells derived from these mice are inferior in serial or competitive transplantation assays.[41-43] Taken together, these data begin to explain why germline complete or nearly complete LOF of a *GATA2* allele could result in Emberger syndrome with lymphedema, ID and/or myeloid malignancy.

Our functional studies together with published phenotypic and clinical data suggest that DNA binding affinity alone cannot explain the association seen between specific mutations and resultant subtypes of disease. We show for the first time that two of these mutations impact on the ability of GATA2 to interact with a known key myeloid transcription factor binding partner. Specifically, gT354M and gC373R bind PU.1 more tightly. PU.1 is an ETS transcription factor that in myelopoiesis induces expression of genes pivotal to myeloid differentiation.[44] Increased binding may sequester both GATA2 and PU.1 from their normal roles in myelopoiesis. Thus, our results provide a clue to how some GATA2 mutants could directly regulate PU.1 protein by affecting its transactivation activity. Since many other myeloid transcription factors interact with the ZF2 domain, we hypothesize that some GATA2 mutants may impair hematopoiesis via aberrant sequestration of key hematopoietic transcription factors resulting in disruption of normal cellular processes such as differentiation, proliferation and survival. Further, gT354M and gR396Q were recently also reported to attenuate chromatin occupancy, reduce endogenous target gene activation and disrupt endogenous GATA2 WT binding

suggestive of DN properties.[38,45] Further investigations to explore GATA2 mutant-mediated alteration in chromatin occupancy, global gene expression profiles and protein-protein interaction studies are warranted.

Previously, we demonstrated that GATA2 mutants inhibited the potent all-*trans* retinoic acid-induced differentiation of the HL-60 myeloid cell line.[2] Here we extend that work to show that GATA2 WT or mutants have varying effects on growth and differentiation of hematopoietic progenitors. Expression of GATA2 WT results in a profound block in clonogenic growth, colony numbers and differentiation *in vitro*. This is consistent with elevated GATA2 levels in hematopoietic stem cells (HSC) driving quiescence[46,47] and the role of GATA2 in maintaining the "stemness" of these cells.[48,49] Expression of GATA2 is known to perturb normal hematopoiesis and cause pancytopenia in transplanted mice.[39] All GATA2 mutants, independent of DNA binding affinity, displayed loss of progenitor growth inhibition. Hence GATA2 mutants may be unable to restrain growth/cycling in the hematopoietic compartment leading to active cycling of HSC, ultimately leading to stem cell pool depletion and exhaustion. This may be particularly important on a background of persistent infection or hematopoietic stresses such as bone marrow failure. This model has been proposed previously for MonoMAC and DCML deficiency syndromes.[6-9] Colony formation assays also show a skewing of monocyte progenitors to granulocytic progenitors following expression of WT and mutants, consistent with these mutants being partial LOF. This skewing in differentiation leading to monocytopenia is observed in MonoMAC and DCML deficiency syndromes. On the other hand, erythroid progenitors are selectively resistant to the suppressive effects of WT, consistent with a previous study in which enforced GATA2 expression promotes expansion of immature erythroid blasts with impaired differentiation.[50] Interestingly, sL359V was the only mutant to maintain the same ability as WT to generate erythroid lineage precursors. While retaining certain

functional properties of the WT protein, sL359V simultaneously up-regulates transcription for some genes and down-regulates for other genes,[2] suggesting that sL359V is a "change-of-spectrum" mutation, a term initially introduced to describe p53 mutants with differential changes in transactivation.[51,52] This is consistent with our functional data in which sL359V demonstrates mild GOF in DNA binding affinity and transactivation in some contexts, but specific LOF in colony formation. Together, our functional assays demonstrate mutation-specific phenotypic characteristics that begin to explain clinical findings.

Recent studies have shown that a subset of myeloid malignancy patients (32%, 16/50) with germline GATA2 mutations including two cases of gT354M and three cases of gR398W also harboured acquired LOF ASXL1 heterozygous mutations.[26,32] ASXL1 is a component of DNA- and/or histone-modifying complexes[53] that is recurrently mutated in various myeloid malignancies.[54-56] The significant co-occurrence of *GATA2* and *ASXL1* mutations strongly suggests that *ASXL1* is an important co-operating genetic event for overt myeloid malignancies manifestation.[57] The concept of "genetic predestination" has received recent support in somatic cases of MDS where initiator or founder mutations in a particular gene direct or constrain the impact of subsequent mutations and phenotypes.[58,59] Here we demonstrate *in vitro* functional differences of various clinically important GATA2 ZF2 mutations and correlate clinical phenotypes constrained by several of these mutations. Additionally, even within the haematological malignancies, three recurrent germline GATA2 mutations (*i.e.* gT345M, gR396Q and gR398W) appear to predominate according to disease subtypes (MDS/AML or CMML), and exhibit clear different clinical outcomes and survival. Hence, we propose that there are GATA2 germline and somatic-, domain- and amino acid residue-specific mutations that impart clonal evolutionary constraints on myeloid malignancy development such that any loss of

DNA binding ability predisposes to MDS/AML while specific mutations have the ability to impact on final subtype within this group of diseases possibly via altered protein-protein interactions.

In contrast to the predominance of *GATA2* germline mutations identified in ZF2, the majority of somatic *GATA2* mutations cluster within ZF1. GATA2 ZF1 mutations and a sole ZF2 (sR362Q) mutation are often associated with AML harbouring compound heterozygous *CEBPA* mutations (bi*CEBPA* AML; ~25% contain ZF1 or sR362Q mutations).[19-21,34,35] This association appears to be less for all other germline ZF2 mutations which seem rarely to co-occur with *CEBPA* mutations.[19,20,34,35] The reason for this is not clear, but implies differences in the leukemogenic mechanisms employed by ZF1 and ZF2 mutations. Alternatively, it may be that different initiator mutations in epigenetic regulator genes such as *ASXL1*, *TET2*, *IDH1*/*2* or *DNMT3A*[60] or pre-existing mutations in other genes such as *CEBPA* generate permissive pre-leukemic environments enabling selective forces for ZF1 or particular ZF2 mutations. Germline ZF1 missense mutations have not been reported, presumably because they are detrimental during embryonic development. We would predict the same is true for sR362Q which is the only recurrent somatic mutation in ZF2 (excluding sL359V) that "behaves" like ZF1 mutations. Various substitutions of R362 (sR362Q/P/G) have been reported in CN pediatric AML[34,35] and adult bi*CEBPA* AML,[19,20] highlighting the importance of changes to this DNA-binding residue in leukemia development.

While biallelic alterations (one germline and one somatically acquired) in *RUNX1*[61,62] and *CEBPA*[63,64] are seen in CN-AML, single heterozygous germline mutations with autosomal dominant inheritance are prevalent in *GATA2*. To date, only 5 cases of biallelic *GATA2* mutations (mostly present in the background of mutated CEBPA) have

been reported.[19-21,65] Given the rarity of biallelic *GATA2* mutations, GATA2 does not fit into the classical LOF tumour suppressor gene model, but seems to better fit a haploinsufficiency model. A likely explanation is that in AML stem cells, a threshold level of residual GATA2 activity is absolutely essential for leukemic cell survival or selection. Furthermore, studies have shown that germline mutations in intron 4 of *GATA2*[66] and in the 5' UTR contain *cis*-acting auto regulatory elements,[13] resulting in hypomorphic expression of GATA2 and predisposition to aplastic anemia, ID and MDS/AML. Thus, a growing body of evidence points to a role for reduced GATA2 activity in leukemic predisposition whether by mutation, gene deletion or reduced expression of normal GATA2 activity.

In conclusion, the role of GATA2 in disease is an intriguing and complex one where domain-specific and amino acid-specific mutations together with environmental stresses on the hematopoietic and lymphatic systems contribute to various myeloid malignancies, immunodeficiencies with a unique spectrum of infections and/or lymphedema. A better understanding of the commonalities and differences of GATA2 mutations will be crucial in determining those molecular interactions with DNA and other proteins that "drive" phenotypic changes from incidental "passenger" changes, ultimately expediting development of not only GATA2-specific, but mutation-specific targeted therapies.

Acknowledgements

This work was supported by grant funding from the National Health and Medical Research Council, Australia (APP1002317, APP1023059), Cancer Council of South Australia (APP626953) and Adelaide Scholarships International from the University of Adelaide. This work has been submitted in partial fulfillment of the requirement for a PhD degree at the University of Adelaide for C-E.C. and P.V.

Authorship contributions

C-E.C., C.N.H. and H.S.S. wrote the manuscript. C.N.H., A.L.B., M. K-H., J.M.M., R.J.D. and H.S.S. planned experiments. C-E.C., P.V., P.H.S., Y.K.L., P.J.B., M.B. and G.A.E. performed functional experiments. S.W.L. provided expert advice on *ex vivo* experimentation and D.T.O.Y. on clinical aspects of data analysis. All authors critically reviewed and approved the manuscript.

Disclosure of conflicts of interest

The authors declare no conflict of interest.

References

1. Zhang SJ, Ma LY, Huang QH, 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):2076-2081.
2. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*. 2011;43(10):1012-1017.
3. Kazenwadel J, Secker GA, Liu YJ, 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):1283-1291.
4. Mansour S, Connell F, Steward C, et al. Emberger syndrome-primary lymphedema with myelodysplasia: report of seven new cases. *Am J Med Genet A*. 2010;152A(9):2287-2296.
5. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet*. 2011;43(10):929-931.
6. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011;118(10):2653-2655.
7. Vinh DC, Patel SY, Uzel G, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood*. 2010;115(8):1519-1529.
8. Bigley V, Haniffa M, Doulatov S, et al. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J Exp Med*. 2011;208(2):227-234.
9. Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*. 2011;118(10):2656-2658.
10. Pasquet M, Bellanne-Chantelot C, Tavitian S, 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):822-829.

11. Mace EM, Hsu AP, Monaco-Shawver L, et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset. *Blood*. 2013;121(14):2669-2677.
12. Ganapathi KA, Townsley DM, Hsu AP, et al. GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia. *Blood*. 2015;125(1):56-70.
13. Townsley DM, Hsu A, Dumitriu B, Holland SM, Young NS. Regulatory Mutations in GATA2 Associated with Aplastic Anemia. *54th American Society of Hematology Annual Meeting Atlanta, GA* 2012:(Abstract 3488).
14. Gao J, Gentzler RD, Timms AE, et al. Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. *J Hematol Oncol*. 2014;7(1):36.
15. Hahn CN, Brautigan PJ, Chong CE, et al. Characterisation of a compound in-cis GATA2 germline mutation in a pedigree presenting with myelodysplastic syndrome/acute myeloid leukemia with concurrent thrombocytopenia. *Leukemia*. 2015;29(8):1795-1797.
16. Dickinson RE, Milne P, Jardine L, et al. The evolution of cellular deficiency in GATA2 mutation. *Blood*. 2014;123(6):863-874.
17. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*. 2014;123(6):809-821.
18. Yan XJ, Xu J, Gu ZH, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet*. 2011;43(4):309-315.
19. Fasan A, Eder C, Haferlach C, 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):482-485.
20. Green CL, Tawana K, Hills RK, et al. GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. *Br J Haematol*. 2013;161(5):701-705.
21. Greif PA, Dufour A, Konstandin NP, 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):395-403.
22. Marceau-Renaut A, Guihard S, Castaigne S, Dombret H, Preudhomme C, Cheok M. Classification of CEBPA mutated acute myeloid leukemia by GATA2 mutations. *Am J Hematol*. 2015;90(5):E93-94.
23. Collin M, Dickinson R, Bigley V. Haematopoietic and immune defects associated with GATA2 mutation. *Br J Haematol*. 2015;169(2):173-187.
24. Wilkinson-White L, Gamsjaeger R, Dastmalchi S, et al. Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. *Proc Natl Acad Sci U S A*. 2011;108(35):14443-14448.
25. Liew CK, Simpson RJ, Kwan AH, et al. Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. *Proc Natl Acad Sci U S A*. 2005;102(3):583-588.
26. Bodor C, Renneville A, Smith M, 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):890-894.
27. Cuellar-Rodriguez J, Gea-Banacloche J, Freeman AF, et al. Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency. *Blood*. 2011;118(13):3715-3720.
28. Holme H, Hossain U, Kirwan M, Walne A, Vulliamy T, Dokal I. Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *Br J Haematol*. 2012;158(2):242-248.

29. Hsu AP, Johnson KD, Falcone EL, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood*. 2013;121(19):3830-3837, S3831-3837.
30. Ishida H, Imai K, Honma K, 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):1273-1276.
31. Kaur J, Catovsky D, Valdimarsson H, Jensson O, Spiers AS. Familial acute myeloid leukaemia with acquired Pelger-Huet anomaly and aneuploidy of C group. *Br Med J*. 1972;4(5836):327-331.
32. West RR, Hsu AP, Holland SM, Cuellar-Rodriguez J, Hickstein DD. Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. *Haematologica*. 2014;99(2):276-281.
33. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999;94(446):496-509.
34. Luesink M, Hollink IH, van der Velden VH, et al. High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood*. 2012;120(10):2064-2075.
35. Shiba N, Funato M, Ohki K, et al. Mutations of the GATA2 and CEBPA genes in paediatric acute myeloid leukaemia. *Br J Haematol*. 2014;164(1):142-145.
36. Kazenwadel J, Betterman KL, Chong CE, et al. GATA2 is required for lymphatic vessel valve development and maintenance. *J Clin Invest*. 2015;125(8):2979-2994.
37. Chen Y, Bates DL, Dey R, et al. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep*. 2012;2(5):1197-1206.
38. Cortes-Lavaud X, Landecho MF, Maicas M, et al. GATA2 germline mutations impair GATA2 transcription, causing haploinsufficiency: functional analysis of the p.Arg396Gln mutation. *J Immunol*. 2015;194(5):2190-2198.
39. Persons DA, Allay JA, Allay ER, et al. Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood*. 1999;93(2):488-499.
40. Ali A, Christie PT, Grigorieva IV, 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):265-275.
41. Ling KW, Ottersbach K, van Hamburg JP, et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med*. 2004;200(7):871-882.
42. Rodrigues NP, Boyd AS, Fugazza C, et al. GATA-2 regulates granulocyte-macrophage progenitor cell function. *Blood*. 2008;112(13):4862-4873.
43. Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*. 2005;106(2):477-484.
44. Nerlov C, Graf T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev*. 1998;12(15):2403-2412.
45. Katsumura KR, Yang C, Boyer ME, Li L, Bresnick EH. Molecular basis of crosstalk between oncogenic Ras and the master regulator of hematopoiesis GATA-2. *EMBO Rep*. 2014;15(9):938-947.
46. Venezia TA, Merchant AA, Ramos CA, et al. Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol*. 2004;2(10):e301.
47. Tipping AJ, Pina C, Castor A, et al. High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood*. 2009;113(12):2661-2672.

48. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science*. 2002;298(5593):597-600.
49. Snow JW, Trowbridge JJ, Johnson KD, et al. Context-dependent function of "GATA switch" sites in vivo. *Blood*. 2011;117(18):4769-4772.
50. Briegel K, Lim KC, Plank C, Beug H, Engel JD, Zenke M. Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. *Genes Dev*. 1993;7(6):1097-1109.
51. Menendez D, Inga A, Resnick MA. Potentiating the p53 network. *Discov Med*. 2010;10(50):94-100.
52. Resnick MA, Inga A. Functional mutants of the sequence-specific transcription factor p53 and implications for master genes of diversity. *Proc Natl Acad Sci U S A*. 2003;100(17):9934-9939.
53. Fisher CL, Randazzo F, Humphries RK, Brock HW. Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family. *Gene*. 2006;369:109-118.
54. Boultwood J, Perry J, Pellagatti A, et al. Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia*. 2010;24(5):1062-1065.
55. Carbuccia N, Murati A, Trouplin V, et al. Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia*. 2009;23(11):2183-2186.
56. Gelsi-Boyer V, Trouplin V, Adelaide J, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol*. 2009;145(6):788-800.
57. Micol JB, Abdel-Wahab O. Collaborating constitutive and somatic genetic events in myeloid malignancies: ASXL1 mutations in patients with germline GATA2 mutations. *Haematologica*. 2014;99(2):201-203.
58. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. *Blood*. 2013;122(25):4021-4034.
59. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-3627; quiz 3699.
60. Grossmann V, Haferlach C, Nadarajah N, et al. CEBPA double-mutated acute myeloid leukaemia harbours concomitant molecular mutations in 76.8% of cases with TET2 and GATA2 alterations impacting prognosis. *Br J Haematol*. 2013;161(5):649-658.
61. Osato M. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. *Oncogene*. 2004;23(24):4284-4296.
62. Preudhomme C, Renneville A, Bourdon V, et al. High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. *Blood*. 2009;113(22):5583-5587.
63. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. *N Engl J Med*. 2004;351(23):2403-2407.
64. Dufour A, Schneider F, Metzeler KH, et al. Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *J Clin Oncol*. 2010;28(4):570-577.
65. Stieglitz E, Liu YL, Emanuel PD, et al. Mutations in GATA2 are rare in juvenile myelomonocytic leukemia. *Blood*. 2014;123(9):1426-1427.
66. Johnson KD, Hsu AP, Ryu MJ, et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J Clin Invest*. 2012;122(10):3692-3704.

Figure Legend

Figure 1. Structural model of GATA2 ZF2 domain bound to DNA and Western blot-electromobility shift assay (WEMSA) of GATA2 WT and mutants. (A) Schematic diagram of GATA2 ZF2 domain bound to DNA showing the location of amino acid residues that are mutated in MDS/AML and/or ID, and Emberger syndrome. Full length *m*GATA2 WT or mutant proteins were expressed in HEK293 cells. Nuclear extract (10μg) was run on a Western blot (B) or allowed to bind to short oligonucleotide probes corresponding to (C) human *GM-CSF* enhancer and (D) human *TCRD* enhancer and WEMSA performed. Oligonucleotide probes with mutated GATA binding sites (Mut) for both *hGM-CSF* and *hTCRD* enhancer probes were used as negative controls. All blots were probed with anti-GATA2 antibody.

Figure 2. GATA2 mutants differ in their DNA binding affinity when quantified by electromobility shift assay (**EMSA) and isothermal titration calorimetry (ITC).** *h*GATA2 ZF2 WT and mutant constructs were expressed and purified, and the effect of the mutation on their DNA-binding affinity was assessed. The ZF2 mutants were first compared using EMSA (*left*) where the ZF2 domains were titrated onto an oligonucleotide derived from the *hGM-CSF* enhancer, labelled with either ^{32}P (A) or fluorescein (B). The average dissociation constant (K_d) for the *hGM-CSF* oligonucleotide was then determined using ITC (*centre*; average ± S.E.M where more than two titrations were conducted; see also Supplemental Figure 5). Dissociation constants (K_d) for WT and mutants (Mut) were then compared as a ratio to demonstrate magnitude of loss of DNA binding affinity (*right*). ND: not determined. K_d could not be obtained using ITC for two mutants (data not shown); gR361L because the observable binding was too weak to be practicably measured using this technique, and gC373R because the protein displayed

abnormal behaviour upon dilution in control experiments that obscured any possible measurement of DNA binding.

Figure 3. GATA2 mutants display variable loss of transactivation capacity on known GATA2 responsive elements. (A) Transactivation of *CD34* enhancer by *h*GATA2 WT and mutants. HEK293 cells were co-transfected with the *GATA2*-responsive *CD34* enhancer elements linked to a LUC reporter and *GATA2* (WT or mutants) expression constructs or pCMV6-XL6 empty vector (EV). (B) Transactivation of *LYL1* promoter by *h*GATA2 WT and mutants. Cos-7 cells were co-transfected using *LYL1* promoter LUC as reporter and *GATA2* (WT or mutants) expression constructs. (C) Transactivation of *CSF1R* promoter by *h*GATA2 WT and mutants. *GATA2* expression constructs were co-transfected with *CSF1R* promoter LUC as reporter into Cos-7 cells (white columns). The synergistic effect of GATA2 with PU.1 on the *CSF1R* promoter was investigated by co-transfecting *GATA2* WT or mutant constructs together with of PU.1 (grey columns). The effect of GATA2 mutants on GATA2 WT was ascertained by co-transfecting GATA2 WT and mutant constructs (1:1) together with PU.1 (black columns). Luciferase activity assays were performed following all transfections and results reported as fold change compared to EV±S.E.M. (n = 4) independent biological replicates. * indicates $p<0.05$ and ** indicates $p<0.01$.

Figure 4. GATA2 mutants alter protein-protein interaction. Whole cell lysates of HEK293 cells co-transfected with expression plasmids for FLAG-tagged GATA2 (WT and mutants) and MYC-tagged PU.1 or FLAG-tagged PU.1 were immunoprecipitated with (A) anti-FLAG antibody, or (B) anti-PU.1 antibody cross-linked to protein A/G agarose. 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. gT354M and gC373R (indicated by *) demonstrate altered binding partner activity.

Figure 5. GATA2 mutants differentially impact clonogenic expansion and haematopoietic differentiation. Murine stem cell virus (MSCV) based retroviral vector expressing *Gata2* WT or mutants (pMSCV-*mGata2*-IRES-GFP) were transduced into LSK cells. GFP expressed cells were sorted 2 days after transduction and immediately plated in methylcellulose medium supplemented with myeloid and erythroid growth factors. Colonies were enumerated and typed on day 7 of culture. (A) Total colony numbers. The number of colonies for LSK cells transduced with vector, GATA2 WT or mutants were counted. GATA2 WT was used as a comparator for statistical analyses. (B-E) Colonies were typed by their morphology and the numbers of each colony type were reported as percentage (see also Supplemental Figure 9). EV served as a comparator for the statistical analyses; (B) CFU-M, (C) CFU-G, (D) BFU-E and (E) CFU-GM/GEMM. All results are plotted as mean ± S.E.M. and n = independent biological replicates for A-E. * indicates $p<0.05$ and ** indicates $p<0.01$.

Figure 6. Different germline *GATA2* mutations are associated with different clinical manifestations and latency. (A) Disease spectrum for 96 individuals carrying gT354M (n = 59), gR396Q (n = 20) or gR398W (n = 17) predisposing mutations. Time to event analyses were performed using the cumulative incidence function with adjustment for competing risk as per Fine and Gray. (B) Cumulative incidence of MDS/AML stratified by GATA2 mutation subtype. Disorders affecting the myeloid compartment displayed high penetrance (85/96, ~89%) in patients who carried *GATA2* mutations. However, the age of onset was significantly different: gR396Q patients had the earliest onset of MDS/AML (median age 18 years) versus gT354M and gR398W patients (median 32 and 40 years, respectively). (C) Cumulative incidence of immunodeficiency in the 3 subtypes of GATA2 mutations. Patients with the gR398W mutation often presented with an immunodeficiency syndrome first, but within the same decade also developed a myeloid malignancy (median age 39), as did about 40% of gR396Q patients. (D) Cumulative

incidence of immunodeficiency syndrome in gT354M patients. In this analysis, AML, MDS and death from any cause were considered competing risk events that prevent or mask the development of immunodeficiency. In gT354M patients, MDS/AML was usually the first sign of the clinical syndrome, whilst a subset of them (20%) presented with immunodeficiency first. (E) Kaplan-Meier plot of leukemia transformation free survival for GATA2 mutants. gT354M patients with MDS were more likely to transform to AML with shorter lead time, or presented with AML as the first clinical manifestation of their inherited disorder. All *p values* (log-rank test) are for statistical comparisons performed at the overall global level. TAD = transactivating domain. C = zinc co-ordinating cysteine.

Figure 7. Schematic diagram summarizing the DNA binding affinities of GATA2 mutants and their correlation with clinical phenotypes. GATA2 mutants demonstrated a unique DNA binding-phenotype pattern and were grouped into 3 classes based on their effects on DNA binding affinity. The Class I mutant displayed WT level or enhanced DNA binding affinity relative to WT and is associated with CML-BC. Mutants with partially reduced DNA binding affinity were grouped into Class II. Patients with such mutations were prone to a spectrum of phenotypes including MDS/AML and/or immunodeficiency syndrome (ID). Class III comprised of mutants with severe or complete loss of DNA binding including partial or whole gene deletion (DEL). Patients carrying Class III mutants predisposed to primary lymphedema in addition to MDS/AML and ID.

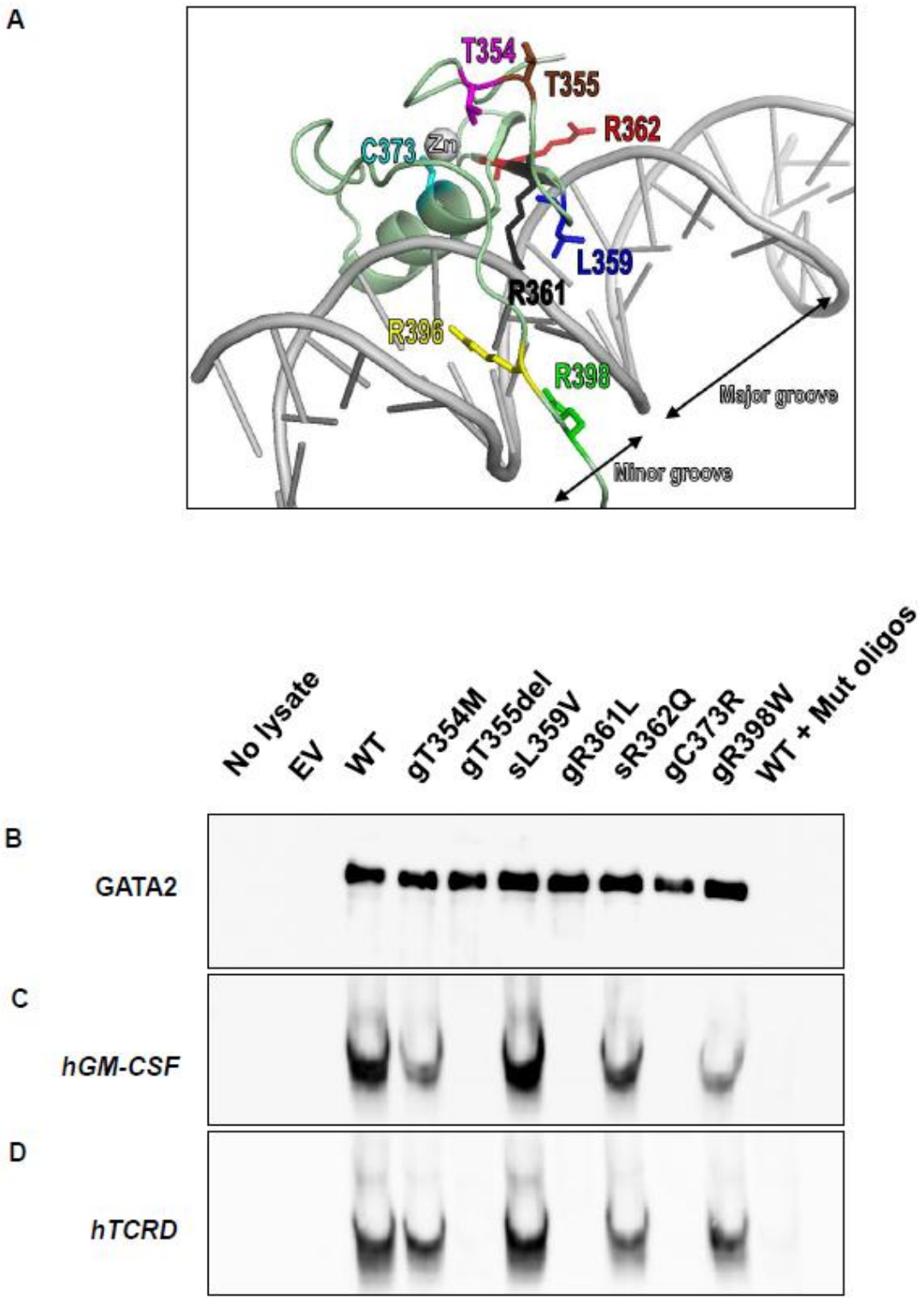

A
T354
T355
R362
C373
Zn
L359
R361
R396
R398
Major groove
Minor groove
No lysate
EV
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W
WT + Mut oligos
B
GATA2
C
hGM-CSF
D
hTCRD

Figure 1

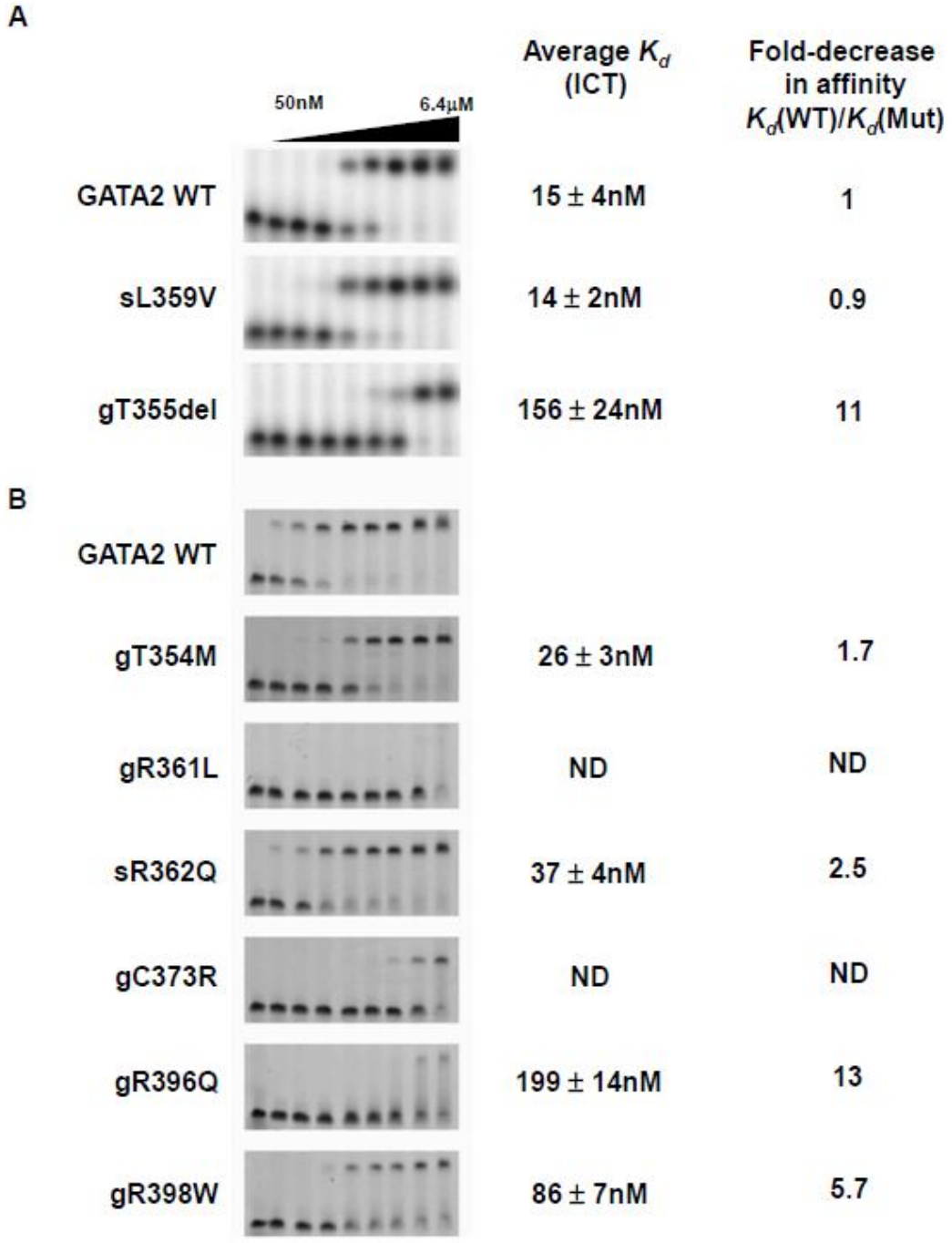

A
50nM
6.4μM
Average K_d (ICT)
Fold-decrease in affinity K_d(WT)/K_d(Mut)
GATA2 WT
15 ± 4nM
1
sL359V
14 ± 2nM
0.9
gT355del
156 ± 24nM
11
B
GATA2 WT
gT354M
26 ± 3nM
1.7
gR361L
ND
ND
sR362Q
37 ± 4nM
2.5
gC373R
ND
ND
gR396Q
199 ± 14nM
13
gR398W
86 ± 7nM
5.7

Figure 2

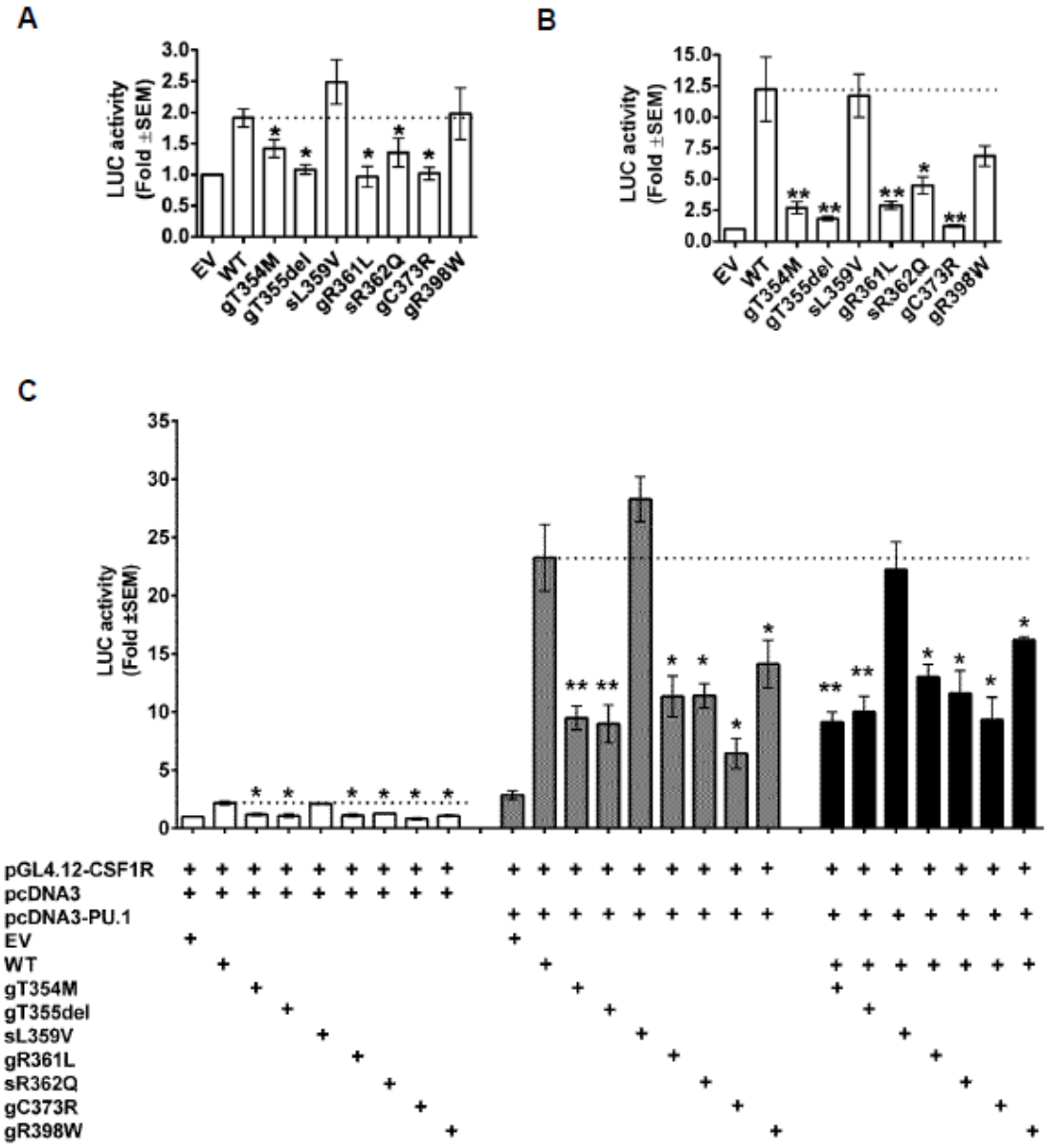

A
LUC activity
(Fold ±SEM)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
EV
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W
B
LUC activity
(Fold ±SEM)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
EV
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W
C
LUC activity
(Fold ±SEM)
35
30
25
20
15
10
5
0
pGL4.12-CSF1R
pcDNA3
pcDNA3-PU.1
EV
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W

Figure 3

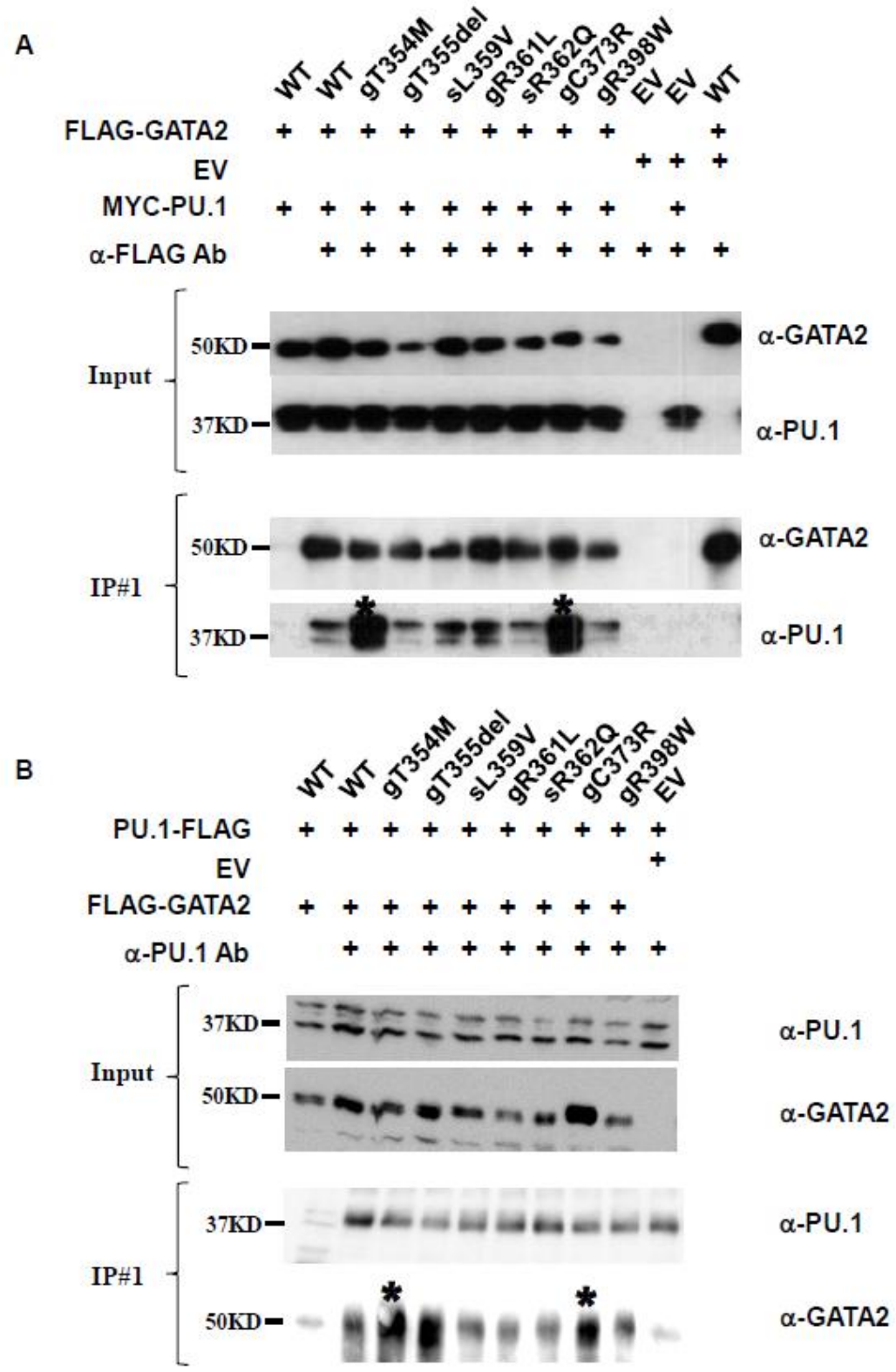

A
WT
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W
EV
EV
WT
FLAG-GATA2
EV
MYC-PU.1
α-FLAG Ab
Input
50KD
37KD
α-GATA2
α-PU.1
IP#1
50KD
37KD
α-GATA2
α-PU.1
B
WT
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W
EV
PU.1-FLAG
EV
FLAG-GATA2
α-PU.1 Ab
Input
37KD
50KD
α-PU.1
α-GATA2
IP#1
37KD
50KD
α-PU.1
α-GATA2

Figure 4

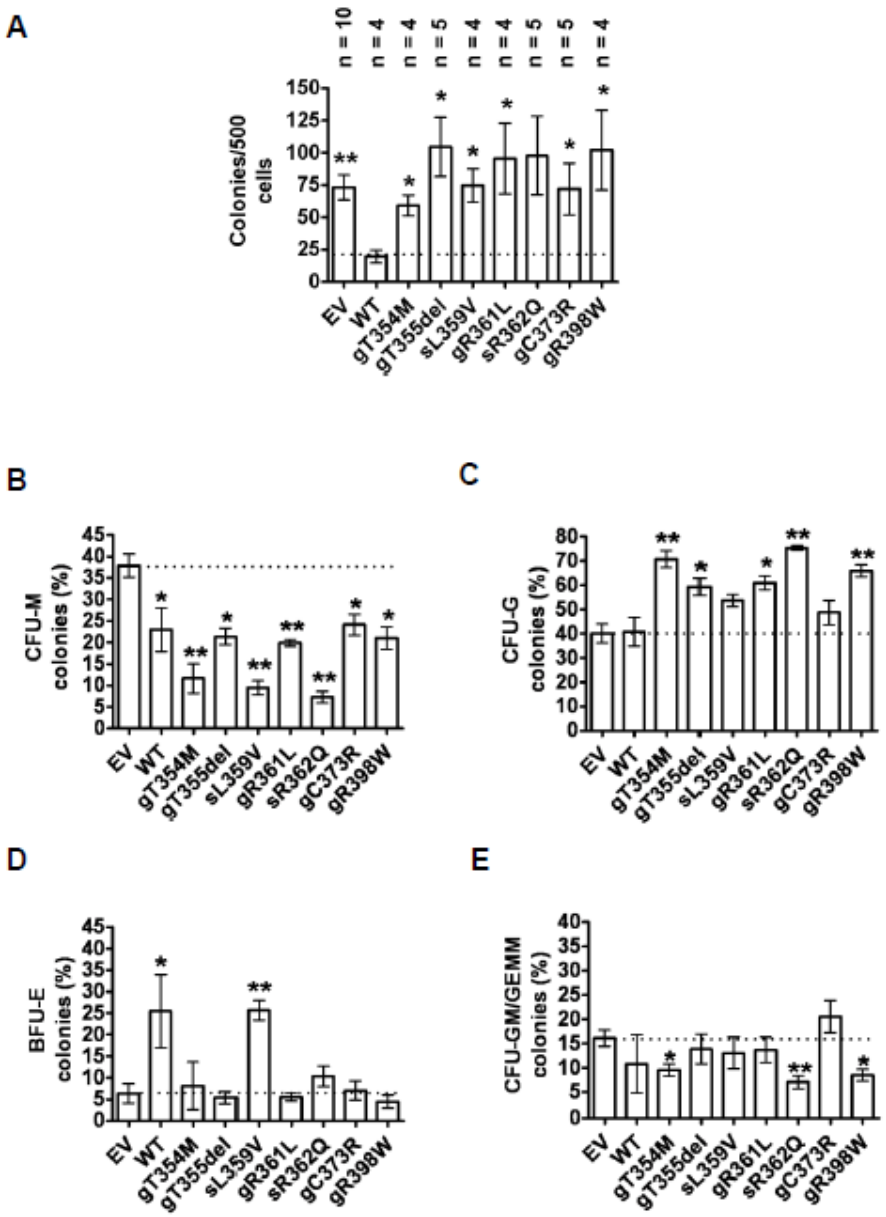

A
n = 10
n = 4
n = 4
n = 5
n = 4
n = 4
n = 5
n = 5
n = 4
Colonies/500 cells
150
125
100
75
50
25
0
EV
WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR398W
B
CFU-M colonies (%)
45
40
35
30
25
20
15
10
5
0
C
CFU-G colonies (%)
80
70
60
50
40
30
20
10
0
D
BFU-E colonies (%)
E
CFU-GM/GEMM colonies (%)

Figure 5

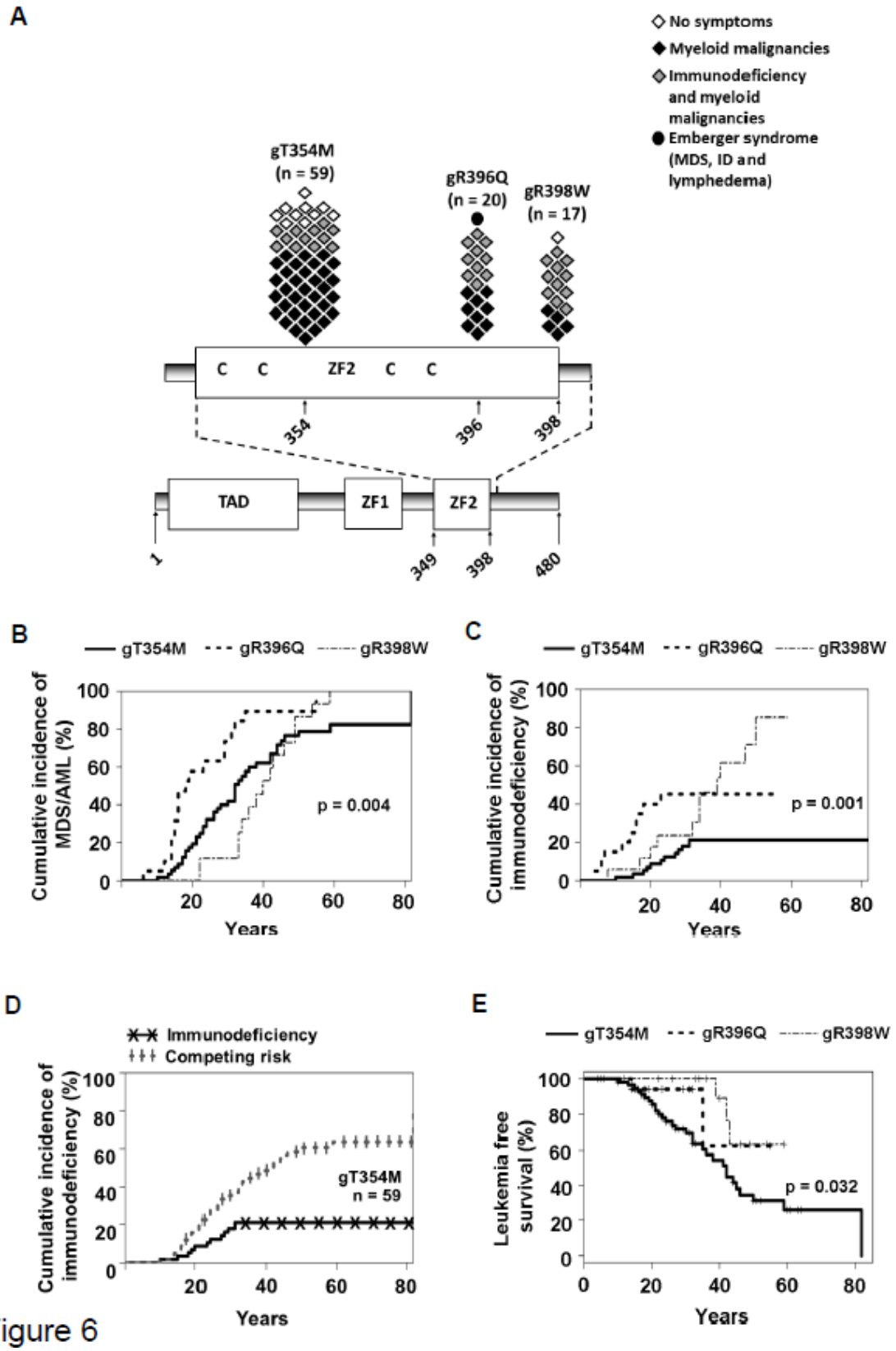

A
No symptoms
Myeloid malignancies
Immunodeficiency and myeloid malignancies
Emberger syndrome (MDS, ID and lymphedema)
gT354M (n = 59)
gR396Q (n = 20)
gR398W (n = 17)
C C ZF2 C C
354
396
398
TAD
ZF1
ZF2
1
349
398
480
B
gT354M gR396Q gR398W
Cumulative incidence of MDS/AML (%)
p = 0.004
Years
C
gT354M gR396Q gR398W
Cumulative incidence of immunodeficiency (%)
p = 0.001
Years
D
Immunodeficiency
Competing risk
Cumulative incidence of immunodeficiency (%)
gT354M
n = 59
Years
E
gT354M gR396Q gR398W
Leukemia free survival (%)
p = 0.032
Years

Figure 6

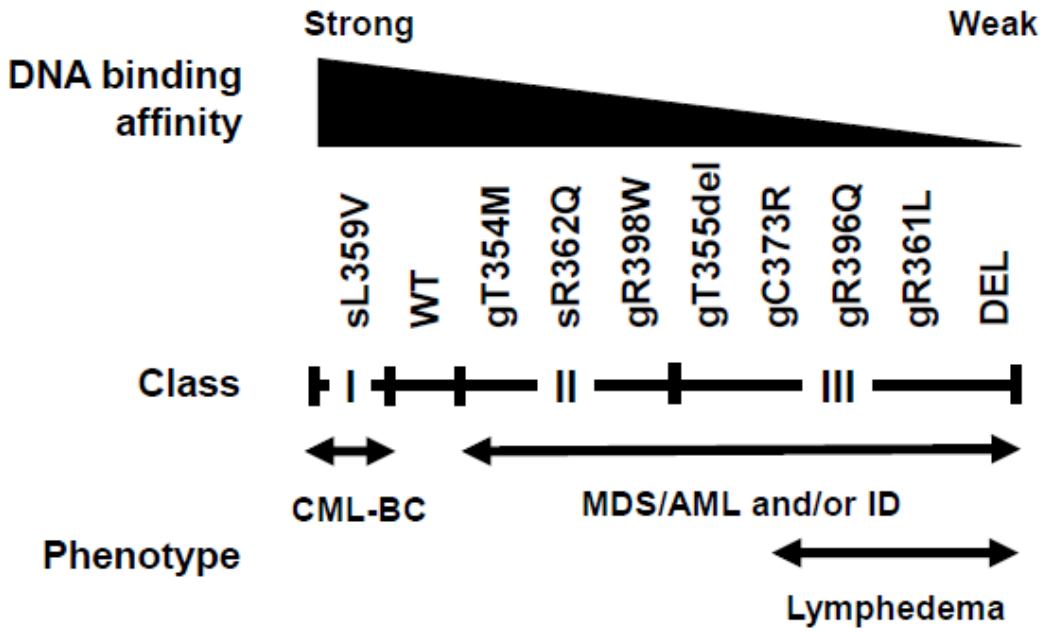

Strong
Weak
DNA binding affinity
sL359V
WT
gT354M
sR362Q
gR398W
gT355del
gC373R
gR396Q
gR361L
DEL
Class
I
II
III
CML-BC
MDS/AML and/or ID
Phenotype
Lymphedema

Figure 7

3.4 Animal model used in the manuscript

The animal model used for the study is a retroviral overexpression model wherein recipient mice were transplanted with LSK cells retrovirally transduced to express WT or mutant GATA2. The ideal model for this experiment would be to generate mice that are heterozygous for the mutation, as this would best represent what is observed in patients. Recently, the emergence of new gene editing technologies like CRISPR/Cas9, TALENs and Zinc finger nucleases have made it easier and quicker to introduce mutations in endogenous gene loci. However, challenges still remain for the generation of knockin mutations (*e.g.* T354M, C373R) which are much less efficient than knocking out the gene of interest. Hence, taking into consideration the time and funding required for this, the retroviral overexpression would still prove to a more feasible approach.

3.5 Effect of *Gata2* haploinsufficiency on mast cell differentiation

Mast cells begin development in the bone marrow where haematopoietic stem cells differentiate into mast cell progenitors through a multi-step process that is tightly regulated by transcription factors. Mast cell progenitors move from the bone marrow and home into tissues where they mature into mast cells. Several transcription factors such as PU.1, STAT5 and GATA2 have been showed to be required for *in vitro* and *in vivo* development of mast cells (93). The interplay between GATA2 and CEBPA has been shown to be critical in lineage determination. Enforced expression of *Gata2* in *Cebpa* deficient myeloid progenitors results in commitment to mast cell lineage (94). High *Gata2* expression is seen in mast cell progenitors, and it has been demonstrated that *Gata2* is required for generation of mast cells (54).

Mast cells are known play a protective role against pathogens like helminths, protozoa, nematodes and certain gram negative bacteria (*Klebsiella pneumoniae*, *Mycoplasma pneumoniae*, *Pseudomonas aeruginosa*, *Haemophilus influenza, E. coli)* (95). They have also been shown to localise to sites of certain viral infections like HIV and HPV, but it is not clear if their role in this context is protective (96). Patients with *Gata2* mutations are reportedly prone to a variety of infections including HPV infections and pneumonia (37).

We hypothesised that haploinsufficiency of *Gata2* would affect mast cell differentiation. We chose to use a mouse model - $Gata2^{+/fl}$ x *Tie2-Cre,* where the *Gata2* allele is floxed out in the haematopoietic and endothelial lineages resulting in haploinsufficiency. We harvested bone marrow cells from $Gata2^{+/+}$ (WT) and $Gata2^{+/-}$ (Het) mice and subjected them to a 6 week mast cell differentiation protocol as previously described (97).

Following this, we assayed for mast cell numbers, secretion of cytokines and degranulation.

Complete blood pictures of donor mice prior to bone marrow harvest revealed normal blood parameters in all mice irrespective of genotype (Table 3.1). Note one wildtype mouse (#90) was slightly neutropenic. The genotypes of mice were confirmed by PCR across the flox cassette on gDNA from bone marrow cells (Figure 3.1).

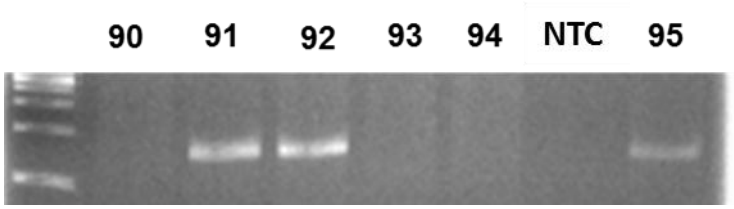


Figure 3.1 Genotypes of mice were confirmed in bone marrow by PCR across the floxed cassette. The 1605 bp product indicates loss of a G*ata*2 allele by Tie2-driven Cre-mediated recombination. No template control – NTC

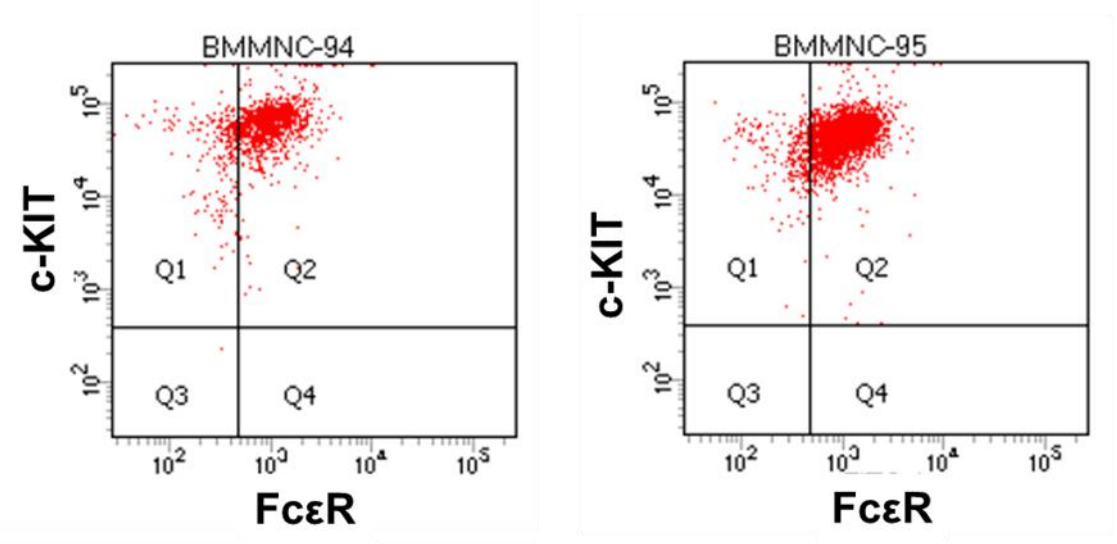


Figure 3.2 Flow cytometry analysis of cell surface expression of c-KIT and FcεR. *In vitro* differentiated mast cells from wildtype (BMMNC-94) and heterozygous (BMMNC-95) G*ata*2 bone marrow mononuclear cells (BMMNCs) show similar numbers of c-KIT and FcεR positivity when analysed by flow cytometry.

Table 3.1. Complete blood pictures of donor mice prior to bone marrow harvest

	Normal range	**90**	**91**	**92**	**93**	**94**	**95**
Genotype (*GATA2*)		+/+	+/-	+/-	+/+	+/+	+/-
White cell count (x 10^9/L)	3.9-13.94	7.8	10.96	9.2	5.92	8.64	6.74
Neutrophils (x 10^9/L)	0.42-2.55	0.29	1.16	1.36	1.01	1.62	1.16
Lymphocytes (x 10^9/L)	2.88-10.98	7.24	9.25	7.64	4.55	6.91	5.53
Monocytes (x 10^9/L)	0.17-0.69	0.27	0.53	0.18	0.2	0.1	0.04
Eosinophils (x 10^9/L)	0.01-0.5	0.01	0.03	0.02	0.11	0.01	0
Basophils (x 10^9/L)	0.00-0.14	0	0	0	0.05	0	0
Neu %		3.72	10.54	14.82	17.1	18.73	17.26
Ly %		92.76	84.38	83	76.78	79.97	82.11
Mo %		3.44	4.83	1.99	3.32	1.17	0.57
Eo %		0.09	0.23	0.17	1.89	0.11	0.03
Ba %		0	0.02	0.03	0.9	0.02	0.03
RBC (x 10^12/L)	7.37-10.9	10.14	10.03	9.45	9.67	9.7	9.08
Haemoglobin (g/L)	109-181	11.2	11.3	10.4	11.2	10.8	10.6
HCT		48.7	53	48.9	50	50	48.7
MCV		48	52.8	51.7	51.7	51.5	53.6
MCH		11	11.3	11	11.6	11.1	11.7
MCHC		23	21.3	21.3	22.4	21.6	21.8
RDW		17.9	16.4	16.3	16.6	17	17.5
Platelets (x 10^9/L)	565-1849	1036	1128	768	637	1126	1090
MPV		5	5.1	4.9	5	4.7	4.9

Table 3.2. Cells recovery following *in vitro* mast cell differentiation of wildtype and *Gata2* heterozygous bone marrow cells.

Mouse ID	***Gata2* Genotype**	**Cells plated**	**Cells recovered**	**Recovered/plated**
90	WT	5.20E+07	6.48E+07	1.25
93	WT	8.90E+07	8.40E+07	0.94
94	WT	9.60E+07	1.13E+08	1.18
95	Het	9.20E+07	1.03E+08	1.12
91	Het	5.20E+07	6.38E+07	1.23
92	Het	5.90E+07	4.56E+07	0.77

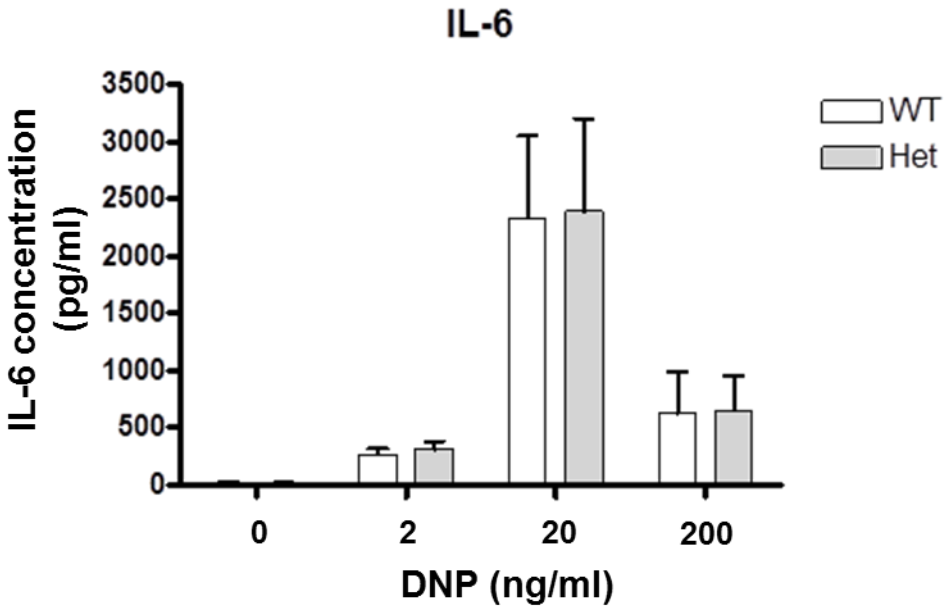


Figure 3.3 Measurement of IL-6 secretion on stimulation with dinitro-phenyl-albumin (DNP) following IgE sensitisation. ELISA for IL-6 revealed no significant differences IL-6 secretion between *Gata2* wildtype and heterozygous bone marrow cells**.**

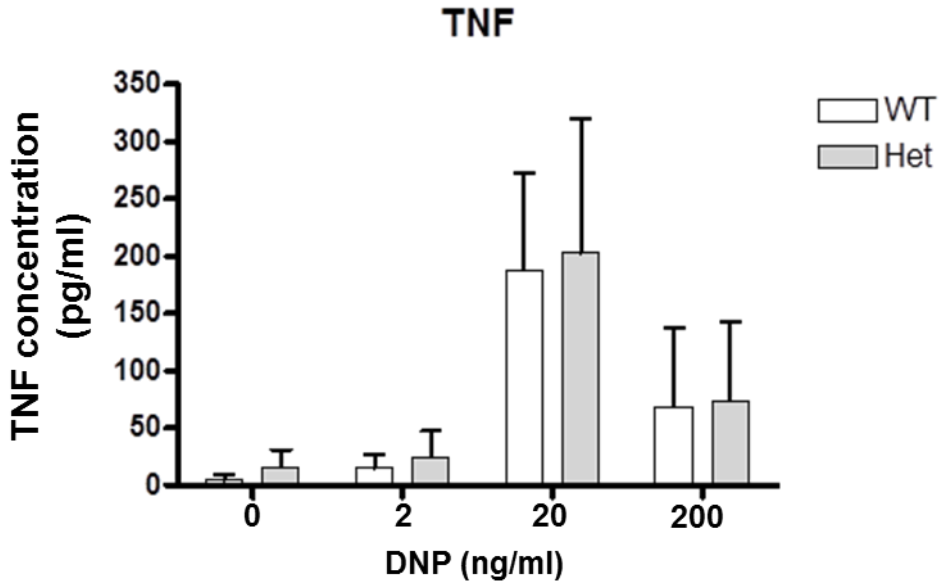


Figure 3.4 Measurement of TNF secretion on stimulation with DNP following IgE sensitisation. ELISA for TNF revealed no significant differences in TNF secretion between *Gata2* wildtype and heterozygous bone marrow cells**.**

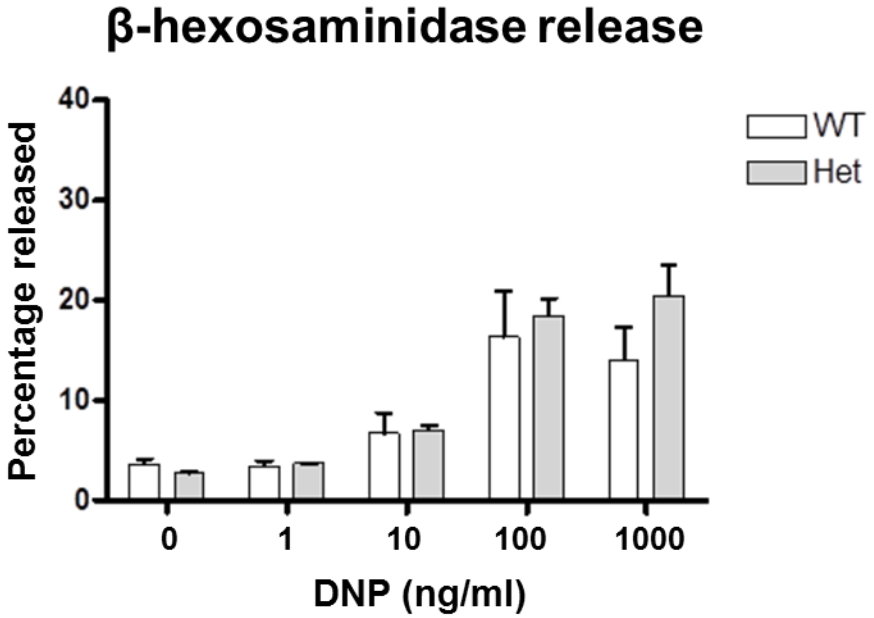


Figure 3.5 Measurement of β-hexosaminidase release. There were no significant differences in mast cell degranulation between wildtype and heterozygous cells.

The percentage of c-Kit^{+}FcεRI^{+} cells was evaluated by flow cytometry and showed no significant differences in these mature mast cell markers (Figure 3.2). There was no change between WT and *Gata2* $^{+/-}$ mice in mast cell numbers (Table 3.2), IL-6 secretion (Figure 3.3) or TNF secretion (Figure 3.4). Degranulation measured by a β-hexosaminidase release assay was also similar (Figure 3.5).

As complete loss of *Gata2* has been shown to have a profound effect on mast cell differentiation, it was surprising that haploinsufficiency of *Gata2* did not have an observable effect in these experiments. In our co-immunoprecipitation assays (discussed in the manuscript in Section 3.3), we observed that specific *GATA2* mutants show altered binding affinities to PU.1 and CEBPA (data not shown). It would be interesting to study the effect of mast cell differentiation in the presence of these mutants given the role of PU.1 and CEBPA in regulating mast cell differentiation. Overall, in the assays that we performed, deletion of one copy of the *Gata2* gene in mice does not perturb differentiation or functioning of mast cells.

Chapter 4: Gene Expression Patterns Improve Existing Prognostic Classification Strategies in Acute Myeloid Leukaemia

4.1. Introduction

Our understanding of the biology of AML has improved tremendously since the advent of genomics technologies like gene expression microarrays and next generation sequencing. Though cytogenetics remains a strong predictor of prognosis in AML, recurrent gene mutations help further classification into prognostically relevant subclasses. Furthermore, by profiling good and poor responders to specific therapeutic approaches, serveral studies have identified gene expression profiles which can be used to predict response (98, 99). The explanation for the heterogeneity seen in AML most likely lies in a combination of cytogenetics, sequence mutations, gene expression, epigenetics, environmental factors and several other factors.

In 2010, an expert international panel representing the European LeukaemiaNet (ELN) published a standardised reporting system for correlation of cytogenetic and molecular genetic data with clinical data (See Table 4.1)(100). The ELN classification system also integrates the World Health Organisation (WHO) classification of AML (See Section 1.2.1.2) which recognises recurrent genetic variations like *NPM1*, *CEBPA* and *FLT3* mutations. It classifies patients into four risk groups – favourable, intermediate 1, intermediate 2 and adverse. Two independent studies have demonstrated that the ELN risk stratification performed well at risk stratification in younger patients, but in older patients, risk stratification was less pronounced as the intermediate groups had very similar survival outcomes(101, 102). The same genetic alteration can have different prognostic significance in patients belonging to different age groups and this could help explain differences seen between older and younger patients within the same risk group(102). Also, differences in treatment strategies for older and younger patients wherein older patients get less intensive treatment, could be yet another factor contributing to this limitation.

Table 4.1 Standardised reporting for correlation of cytogenetic and molecular genetic data in AML with clinical data proposed by European LeukemiaNet (100)

Genetic group	Subsets
Favourable	t(8;21)(q22;q22); *RUNX1-RUNX1T1* inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11* Mutated *NPM1* without *FLT3*-ITD (normal karyotype) Mutated *CEBPA* (normal karyotype)
Intermediate-I*	Mutated *NPM1* and *FLT3*-ITD (normal karyotype) Wild-type *NPM1* and *FLT3*-ITD (normal karyotype) Wild-type *NPM1* without *FLT3*-ITD (normal karyotype)
Intermediate-II	t(9;11)(p22;q23); *MLLT3-MLL* Cytogenetic abnormalities not classified as favourable or adverse†
Adverse	inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1* t(6;9)(p23;q34); *DEK-NUP214* t(v;11)(v;q23); *MLL* rearranged -5 or del(5q); -7; abnl(17p); complex karyotype‡

*Includes all AMLs with normal karyotype except for those included in the favourable subgroup; most of these cases are associated with poor prognosis, but they should be reported separately because of the potential different response to treatment.
†For most abnormalities, adequate numbers have not been studied to draw firm conclusions regarding their prognostic significance.
‡Three or more chromosome abnormalities in the absence of one of the WHO designated recurring translocations or inversions, that is, t(15;17), t(8;21), inv(16) or t(16;16), t(9;11), t(v;11)(v;q23), t(6;9), inv(3) or t(3;3); indicate how many complex karyotype cases have involvement of chromosome arms 5q, 7q, and 17p.

Li *et al*. attempted to address inadequacies of ELN risk stratification by integrating a 24 gene expression signature into it. Their approach reduced the number of risk groups to three and improved resolution between intermediate and adverse outcome groups when compared to ELN risk stratification(103). Assaying expression patterns of 24 genes can be impractical in a diagnostic scenario. Hence we aimed to identify a manageable number of genes with the highest prognostic value in our AML cohort utilizing a high throughput qRT-PCR system which is more accurate and sensitive than microarrays.

We chose 166 AML samples for which we had comprehensive information on mutation status, treatment strategies, cytogenetics, remission, relapse and survival, to study gene expression patterns. We designed a 91 gene panel (see Appendix VI for rationale behind gene selection and AppendixVII for primer sequences) that included genes that are known

to be causative and/or have prognostic value in MDS/AML and transcription factors crucial for cell fate determination in haematopoiesis have also been implicated in AML. *GATA2* is one such gene whose expression is of prognostic value in AML; and has been identified to be mutated in familial MDS/AML as well as sporadic AML. It has an important role in normal and malignant haematopoiesis. Due to our special interest in *GATA2*, we included known binding partners such as CEBPA and SPI1 (PU.1), regulators of *GATA2* and GATA2-target genes in our panel as well to gain a better understanding of how they cooperate/coordinate in leukaemogenesis.

This is a collaborative project in which the samples were provided by Richard D'Andrea, Ian Lewis and Anna Brown; Sonya Diakiw, Grant Engler and Milena Babic were involved in RNA isolation and cDNA synthesis; mutation screening was performed by Bik To, Andrew Wei, Nik Cummings and the Therapeutic Products Facility staff; Fluidigm GE chips were run in conjunction with Milena Babic at the ACRF Cancer Genome Facility; ELN status was determined by Mahmoud Bassal who also helped automate parts of the correlation analyses using R; R pipeline for random forest analysis was customised by Justine Marum; data analysis was supervised and critically reviewed by Chris Hahn, Anna Brown and Hamish Scott.

4.2. Results:

4.2.1 Cohort Characteristics

The median age at diagnosis was 64 years with a range of 18 to 89 years wherein 60% of the patients were male. Only 37% of patients were under the age of 60 years. The distribution of FAB subtypes (See Section 1.2.1 for description) is listed in Table 4.2. Among patients whose ELN risk status was known, 19% were classified in the favourable, 32% in intermiediate-1, 23% in intermediate-2 and 26% in the adverse risk categories.

Table 4.2 Number of cases in each of the FAB subtypes

FAB	Patients
M0	4
M1	36
M2	68
M3	5
M4	29
M5	12
M6	2
M7	1
Unknown	6

4.2.1.1 Mutational Profile

Consistent with previous reports, 45% of the patients had cytogenetically normal AML while 13% had complex karyotypes. *WT1* (32%), *NPM1* (24%), *FLT3* (ITD – 23%, TKD – 7%), *TET2* (21%) and *IDH2* (18%) were the most recurrently mutated genes among those tested (Figure 4.1). Mutations in *IDH1* and *IDH2* were found to be mutually exclusive as previously described (104). Though mutations in *IDH1/2* and *TET2* have been reported to be mutually exclusive in AML, we found *IDH2* mutations in 4/17 (23.5%) individuals with *TET2* mutations while *IDH1* mutations remained mutually exclusive with *IDH2* and *TET2* mutations. Interestingly, *FLT3*-TKD and *IDH1* mutations were only observed in individuals over the age of 55 in our cohort.

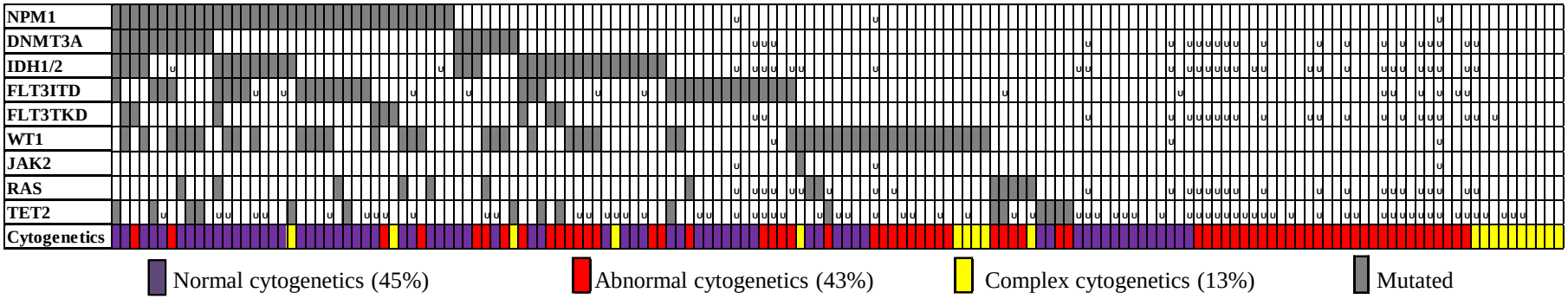


Figure 4.1 Oncoprint showing mutations identified in genes recurrently mutated in AML. Mutations in *DNMT3A, IDH1* and *IDH2* frequently co-occur with *NPM1* mutations. Mutations in *IDH1* and *IDH2* are mutually exclusive. U indicates samples where mutation status is unknown.

4.2.1.2 Quality Control and Data Normalisation

Expression of 95 genes in 166 bone marrow mononuclear samples collected at AML diagnosis, 8 normal CD34 selected bone marrow samples and 13 normal BMMNC control samples were analysed by qRT-PCR using two 96.96 Dynamic Arrays utilising EvaGreen chemistry on the Biomark HD System (Fluidigm), alongside K562 and water controls. *HOXA9* was measured in duplicate for both runs as a reproducibility control. The expression of this gene was remarkably consistent (Pearson's correlation coefficient r = 0.99) between duplicates indicating excellent chip reproducibility. Genes with failed calls or multiple peaks in melting curve analyses across samples and samples that failed across >50% genes were flagged and eliminated from any expression analysis.

Selection of optimal housekeeping genes for normalisation is an important step in data analysis as this can have a strong impact on the way the data is interpreted. Normalisation usually accounts for quality and quantity of RNA, qRT-PCR efficiency and variations across the chips. Using housekeeping genes, which are highly expressed in most tissues, for normalisation used to be common practice but is far from ideal as it tends to skew data especially when the panel of genes to be analysed has a wide range of expression. Therefore, it is recommended to use the geometric mean of expression of at least two housekeeping genes which are most stable across the samples of interest and preferably which have different expression levels when normalising data of widely expressing gene sets. GeNorm analysis ranks genes according to their expression stability which is measured as the average pairwise variation V for that gene with all other tested housekeeping genes (105). We included four housekeeping genes with a wide expression ranges in our panel - *B2M*, *RPLP0*, *HMBS* and *HPRT1*. Raw expression values (calculated as 2^{-Ct}) of the four housekeeping genes from a random selection of samples from both runs was analysed using geNorm to determine the most stable housekeeping genes (Figure 4.2). *RPLP0* was found to be the most stable housekeeping gene whereas *B2M* was the least stable. Hence, each sample was normalised to the geometric mean of expression of *RPLP0*, *HMBS* and *HPRT1*. The study conducted by Vandesompele *et al.* to validate their geNorm algorithm, tested the algorithm on different tissue panels and reported *B2M* to be one of the worst scoring genes in their analyses and *HPRT1* to be among the best. Our data further corroborates this finding, at least for the AML, $CD34^{+}$ and BMMNC cells in this study.

geNorm

100%

geNorm PrimerDesign

Sample ID	B2M	RPLP0	HMBS	HPRT1	Normalisation Factor
80	1.53E-02	2.85E-03	1.75E-05	1.01E-05	0.2224
19	1.62E-02	6.07E-03	8.36E-06	1.58E-05	0.4728
67	1.63E-02	2.14E-03	3.89E-06	1.11E-04	0.1670
16	1.68E-02	5.16E-03	3.35E-05	7.77E-06	0.4018
6	1.77E-02	8.41E-03	6.87E-05	5.14E-05	0.6551
14	1.90E-02	2.31E-02	1.01E-04	1.13E-04	1.7975
9	1.94E-02	2.98E-02	8.98E-05	4.69E-05	2.3209
23	1.96E-02	1.36E-02	3.60E-05	4.31E-05	1.0609
57	2.06E-02	2.92E-03	1.37E-05	6.57E-05	0.2271
52	2.13E-02	6.66E-02	3.27E-04	2.34E-04	5.1858
73	2.18E-02	4.20E-02	1.18E-04	2.37E-04	3.2678
72	2.19E-02	3.43E-02	1.03E-04	3.75E-05	2.6711
96	2.26E-02	2.96E-02	3.17E-05	1.92E-04	2.3056
92	2.29E-02	4.44E-02	7.73E-05	1.90E-04	3.4617
64	2.32E-02	1.02E-01	4.91E-04	6.94E-04	7.9327
84	2.33E-02	5.11E-02	1.12E-04	1.70E-04	3.9769
59	2.45E-02	2.03E-02	2.76E-05	8.10E-05	1.5846
22	2.66E-02	7.66E-02	3.80E-04	1.78E-04	5.9684
49	2.72E-02	3.21E-02	2.04E-04	2.98E-04	2.4993
51	2.84E-02	1.16E-01	2.39E-03	2.79E-03	9.0154
13	2.92E-02	7.33E-02	4.13E-04	3.63E-04	5.7109
58	3.15E-02	1.15E-01	2.82E-03	2.58E-03	8.9446
25	3.29E-02	3.56E-02	1.06E-04	1.78E-04	2.7749
68	3.88E-02	3.66E-02	3.07E-04	1.66E-04	2.8519
62	4.32E-02	5.07E-02	1.12E-03	1.90E-03	3.9507
48	5.51E-02	4.69E-02	1.74E-04	2.72E-04	3.6566
M < 1.5	1.853	1.496	1.728	1.740	

Figure 4.2 Stability of housekeeping genes was assessed using GeNorm analysis. *RPLP0* (highlighted in green) was found to be the most stable housekeeping gene across the samples whereas *B2M* (highlighted red) was the least stable. M – Average pairwise variation for that gene with all other tested reference genes. Normalisation factors are calculated for each sample based on the geometric mean of all the housekeeping genes.

4.2.1.3 Gene Expression patterns associated with *NPM1* mutations

NPM1 (Nucleophosmin 1) encodes a nucleolar phosphoprotein that is the most frequently mutated gene in cytogenetically normal AML. The wildtype protein shuttles between the cytoplasm and nucleus and is known to play a role in ribosome biogenesis, centrosome duplication, chromatin remodelling, transcription and cell proliferation(106). Recurrent mutations replace the nucleolar localisation signal with a

strong nuclear export signal leading to retention of NPM1 in the cytosol. The negative effect of the mutant protein is further intensified by sequestration of the wildtype protein from the nucleolus caused by oligomerisation of wildtype and mutant NPM1(107). In our cohort, *NPM1* mutations were found in 44% of patients with cytogenetically normal AML. Considerable overlap was seen between *NPM1* mutations and *DNMT3A*, *IDH1*, *IDH2* and *TET2* mutations – 57% *NPM1* mutated AML also had a mutation in at least one these genes (Figure 4.1). *NPM1* mutated AMLs have been identified to cluster into four different methylation profiles by Figueroa *et al* (108). This may be explained by the frequent co-occurrence of mutations in these methylation related genes.

When we analysed gene expression of *NPM1* mutated AMLs, we found several genes were significantly up- or down-regulated in these AMLs in comparison to the remaining AMLs (Figure 4.3). We reconfirmed some previously reported correlations including high expression of *MEIS1, HOXA9, PBX3, PLA2G4A* and low expression of *CD34, MN1, TAL1*and *FGFR1* in NPM1 mutated AML (109). We also found previously unknown correlations such as high expression of *CEBPA*, *C1RL* and low expression of *ITPR3, IKZF2, BSPRY* and *PRDX2.*

HOXA9 and *PBX3* are overexpressed consistently in *NPM1*-mutated AML and an overexpression signature of these genes alongside *HOXA11* and *HOXA7* is associated with shorter OS in cytogenetically abnormal AML(103). PBX and MEIS family proteins are the most important HOX co-factors as they help increase both selectivity and specificity of HOX proteins(110). Knockdown of *HOXA/PBX3* renders cells sensitive to chemotherapy. It has been shown that cells overexpressing *HOXA9* and *PBX3* are especially sensitive to HXR9, a drug that prevents formation of *HOX/PBX* heterodimers. Since *NPM1* mutated AML accounts for ~30% of all AMLs, therapeutically targeting *HOXA9/PBX3* may have applications in treating a substantial number of AML patients.

PRDX2 is a member of the peroxiredoxin family that regulates reactive oxygen species and is implicated as a tumour suppressor which is often epigenetically silenced in AML (111). Interestingly, low expression of *PRDX2* was observed in *NPM1* mutated AML and is likely a consequence of specific epigenetic patterns associated with NPM1 mutations (108).

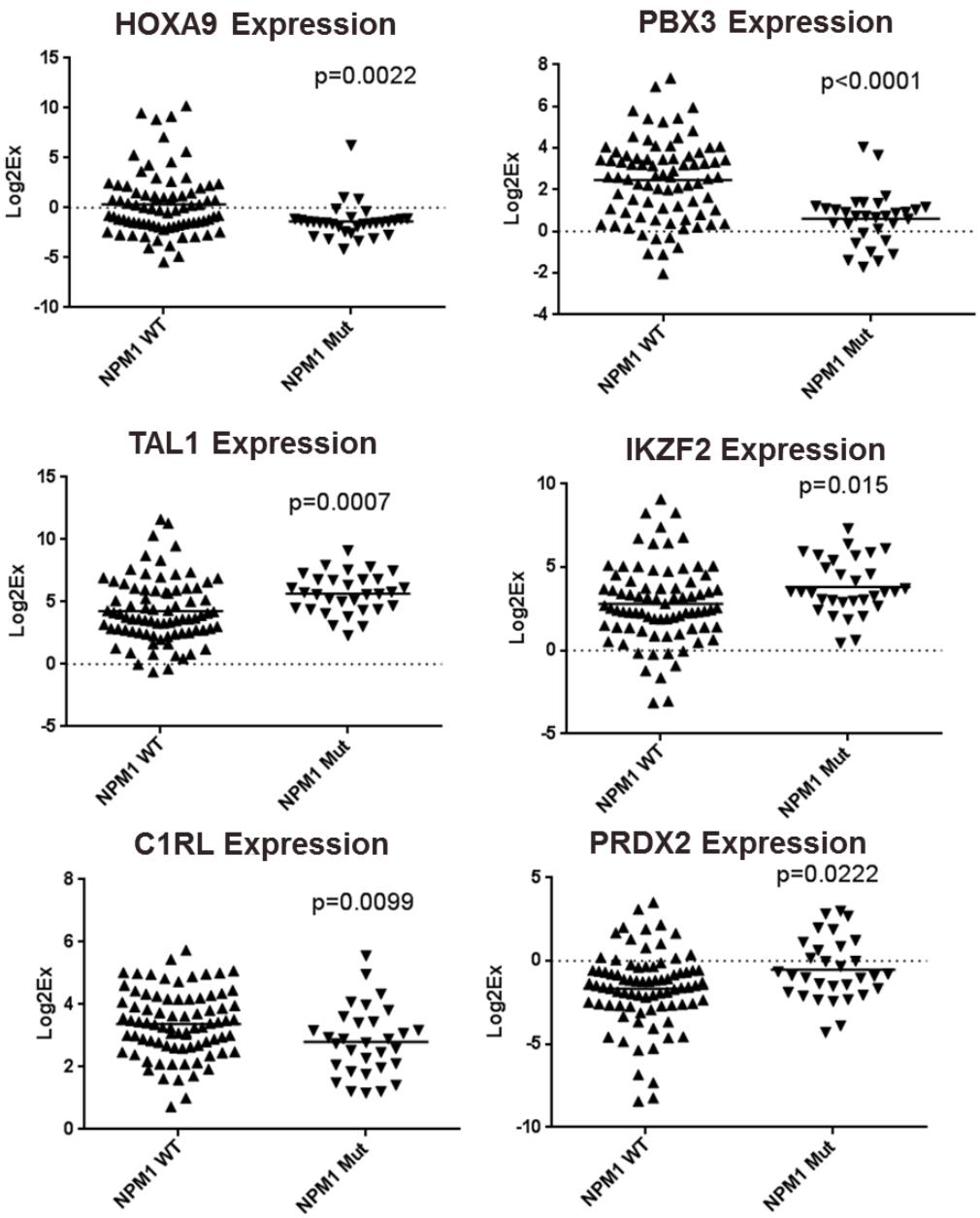


Figure 4.3 Distinct gene expression patterns associated with NPM1 mutations. High expression of *HOXA9, PBX3, C1RL, PRDX2* and low *TAL1* and *IKZF2* were seen in *NPM1*-mutated AML. Delta Ct values are shown on the y-axis.

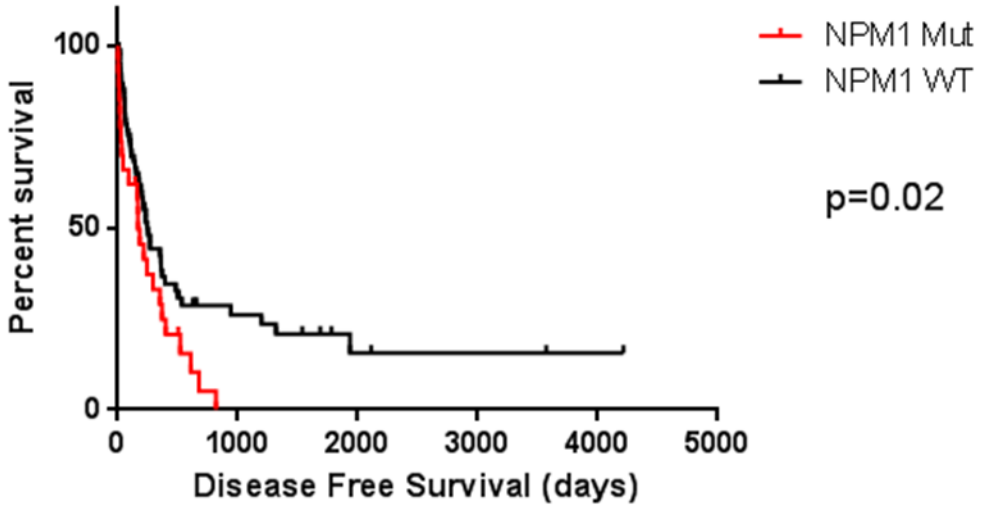


Figure 4.4 Patients with NPM1 mutations have poorer disease free survival than those with wildtype NPM1.

Though *NPM1* mutations are generally associated with favourable prognosis, we found that *NPM1* mutated patients had worse DFS in our patient cohort (Figure 4.4). 14/31 *NPM1* AMLs used for this analysis, also harboured *FLT3* ITD mutations. This may have skewed the data, as the presence and mutation load of *FLT3* ITD reportedly affects the favourable impact of *NPM1* mutations(112). On separating the *FLT3* ITD mutated AML from *FLT3* wildtype AML within the *NPM1* mutated group, there was no obvious difference in survival outcomes (data not shown). Moreover, there are conflicting reports in the literature regarding the effect of *NPM1* and *FLT3* ITD mutations on disease outcome (113-115).

4.2.2 Correlation of Gene Expression Patterns

Correlation of expression between genes was analysed by calculating Pearson's correlation coefficients. The correlation coefficients are represented in a heatmap format generated in R in Figure 4.5.

4.2.2.1 Epigenetic modifier genes are co-regulated across AML

Epigenetic modifier genes are commonly mutated in myeloid malignancies (See Figure 4.1 and Section 1.5). Moreover epimutations (*i.e.* aberrant DNA methylation changes at specific sites in the genome) have been shown to contribute to malignant transformation underlining the importance of epigenetics in AML. For example, frequent hypermethylation in an internal promoter of *DNMT3A* in AML is believed to mimic *DNMT3A* mutations, synergising with *IDH1/2, RUNX1* and *NPM1* mutations to alter

DNA methylation, histone modification, hematopoietic differentiation and cell survival (116).

Expression of chromatin modifier CTCF was highly positively correlated with epigenetic genes *ASXL1* (Pearson's correlation coefficient r = 0.91), *EZH2* (r = 0.91), *SUZ12* (r = 0.92), *MLL* (r = 0.82) and *TET2* (r = 0.74). Though the relevance of this correlation to disease type and prognosis is not understood currently, it suggests that they are co-regulated across AML samples irrespective of subtype or mutation status (Figure 4.5). Intriguingly, mutations in these epigenetic regulators co-occur in myeloid malignancies. For example, *EZH2* and *TET2* mutations are frequently associated with mutations in *ASXL1*(117, 118). Muto *et al.* demonstrated that concurrent loss of *TET2* and *EZH2* cooperate to give an MDS phenotype in mice. These findings suggest that co-expression of these epigenetic regulators may have an additive or synergistic effect. It is also likely that expression of each gene contributes to a different aspect of oncogenicity.

4.2.2.2 *ERG* and *GATA2* expression is positively correlated in AMLs

ERG (ETS related gene) encodes a transcription factor that is essential for definitive haematopoiesis and also plays a role in HSC maintenance. It is frequently rearranged in myeloid malignancies. ERG has also been shown to be a direct upstream regulator of *Gata2* during development (119). High *ERG* expression in AML is associated with poor outcome (120). *ERG* expression is positively correlated (r = 0.56) with *GATA2* expression in AML and BMMNC in our cohort. This confirms previous findings where upregulated *Gata2* was reported in an *Erg* overexpressing mouse leukaemia model (121). While they also reported an activation of RAS pathway with elevated levels of active RAS-GTP and its downstream effectors, in our AML cohort, consistent downregulation of *ERG* was observed in *RAS* mutant AML. This suggests that when RAS signalling is constitutively active, there is no requirement for *ERG* to be upregulated. A recent study by Huang *et al.* showed that MAPK/ERK-mediated phosphorylation of ERG enhances stem cell and leukemic gene expression signatures, promoting HSPC proliferation thereby contributing to poor outcome in leukaemias expressing high ERG (122).

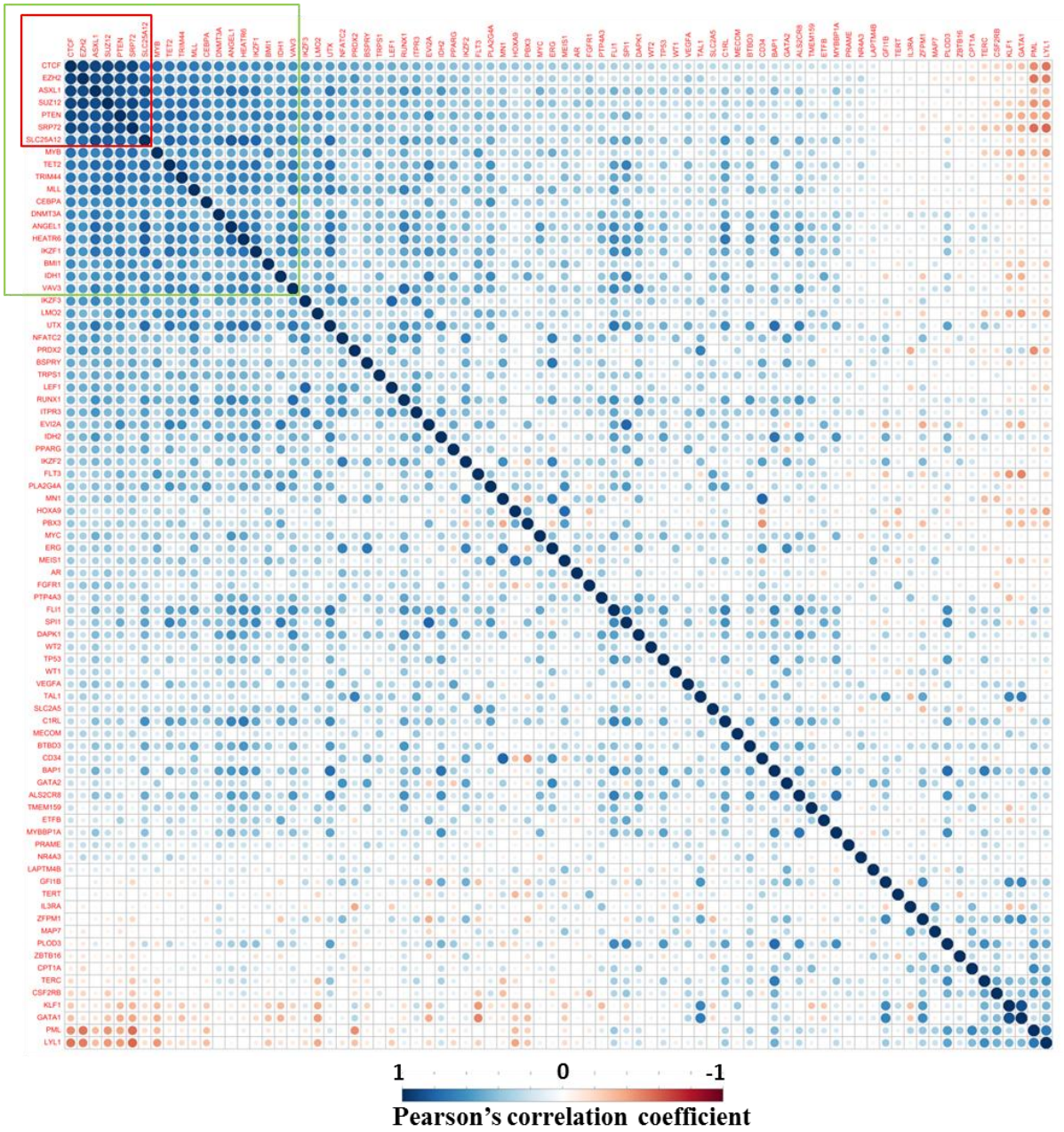


Figure 4.5 Pearson's correlation coefficient of 80 genes vs 80 genes. Visualisation of Pearson's correlation coefficients of every gene against every other gene (excluding housekeeping genes) on the 85 gene panel in a heatmap format. The colours range from dark blue (strong positive correlation) to white (no linear correlation) to red (strong negative correlation).

4.2.2.3 *EVI2A* is consistently upregulated in *RAS* mutant AML

All samples with *RAS* mutations show a consistent upregulation of *EVI2A* expression. The gene encoding EVI2A (Ecotropic Viral Integration site 2A) resides within the intronic region of *NF1*. NF1 is RAS GTPase activating protein (RAS-GAP) which catalyses the conversion of activated RAS-GTP to inactivated RAS-GDP. It seems likely that upregulation of *EVI2A* expression can lead to downregulation of *NF1* via

interference of *NF1* transcription caused by transcriptional upregulation of the nested gene *EVI2A*. It has been demonstrated in a mouse model that oncogenic RAS cooperates with loss of *NF1* to give rise to more aggressive AML-like disease than either of these aberrations individually (123). This cooperation may be a result of the involvement of NF1 in other signalling pathways. It can also be due an additive effect of NF1 loss on the other wildtype RAS isoforms within the cell resulting in further activation of RAS signalling. Another such correlation between mutation of *NRAS/KRAS* and low *GATA2* expression has been followed up and discussed in Chapter 5 (See Section 5.2).

Table 4.3. Glossary

Random Forest: A Random Forest consists of a collection or ensemble of decision trees, which can be used to predict the final outcome. The random forest algorithm was developed by Breiman in 1984.
Recursive partitioning analysis: Techniques to construct subgroups that are as homogeneous as possible internally with regard to outcome and as separated as possible externally (124)
Decision tree: Decision trees are data mining tools which predict either a categorical or continuous response variable using a tree-like model.
VIMP: To calculate the Variable importance measure for each variable included in the tree, the variable values are permuted and the oob error is recalculated, and compared to the original oob error. The increase in oob error is an indication of the variable's importance. The aggregate oob error and importance measures from all trees determine the overall oob error rate and Variable Importance measure. Variables are then ranked based on VIMP, where the largest VIMP is ranked the highest, and deemed the most important for prediction of outcome
Overfitting: Overfitting occurs when a model conforms too closely to the data (especially as most data have at least a small degree of error) and ends up describing random error or noise instead of the actual underlying relationship.

4.2.3. Random Forest Analysis for Survival Prediction

To assess if any genes in our panel were associated with survival outcomes, we applied our dataset consisting of survival and gene expression data of 85 genes in 122 AML patients, to a random forest analysis. Random forest analysis generates decision trees utilising recursive partitioning which is a fundamental tool in unbiased data mining. A

decision tree is a model that predicts the value of a target variable (survival outcome in this case) based on several input variables (gene expression). The R package – rfsrc (random forests for survival, regression and classification) was used to simulate 1000 random classification trees. Nodes are split based on a log-rank splitting rule wherein splits (cut-off for expression data in this case) are determined such that it maximises the log-rank test statistic. In other words, each split in a tree is based on an expression cut-off that best separates the survival outcomes. For each classification tree generated, a particular number of samples are randomly picked as the training set and the remaining are used to validate the tree. A prediction error rate is recorded, based on which variable importance (VIMP) measures are calculated and the genes are ranked based on importance for each gene (See Appendix VIII for full list).

Table 4.4 Top ten genes for each survival outcome

Rank	Overall survival (OS)	Disease Specific Survival (DSS)	Disease Free Survival (DFS)	Event Free Survival (EFS)
1	***BSPRY***	***BSPRY***	***LAPTM4B***	***LAPTM4B***
2	***LAPTM4B***	*RARA*	***BSPRY***	***BSPRY***
3	***IDH1***	***ERG***	*PRAME*	***NR4A3***
4	***NR4A3***	*LAPTM4B*	*RARA*	*ETFB*
5	*ETFB*	*HEATR6*	*ETFB*	*DAPK1*
6	***ERG***	***IDH1***	***NR4A3***	*PRAME*
7	*SPI1*	*LMO2*	*PLA2G4A*	***IDH1***
8	*DAPK1*	***NR4A3***	***ERG***	***ERG***
9	*HEATR6*	*TP53*	***IDH1***	*RARA*
10	*LMO2*	*FLI1*	*LMO2*	*SPI1*

The five genes featuring in the top ten genes for each outcome were selected – *BSPRY*, *LAPTM4B*, *IDH1*, *NR4A3* and *ERG* (highlighted in Table 4.4). A preliminary classification tree with ten terminal nodes (Figure 4.6) was generated by applying the rpart (recursive partitioning and regression trees) package in R to expression data from these five genes. Complex classification trees such as this have a risk of overfitting, which occurs when a model conforms too closely to the data (especially as most data have at least a small degree of error) and ends up describing random error or noise

instead of the actual underlying relationship. Hence, to prevent overfitting the data, this preliminary tree with ten nodes was then pruned to minimise the 10-fold cross validation error (Figure 4.7). This generated a pruned decision tree with four nodes. The mean error rate for the survival outcomes is 43.7%, a comparable error rate to previous reports of validated classification tree models, suggesting that the data has not been over-fit (125, 126).

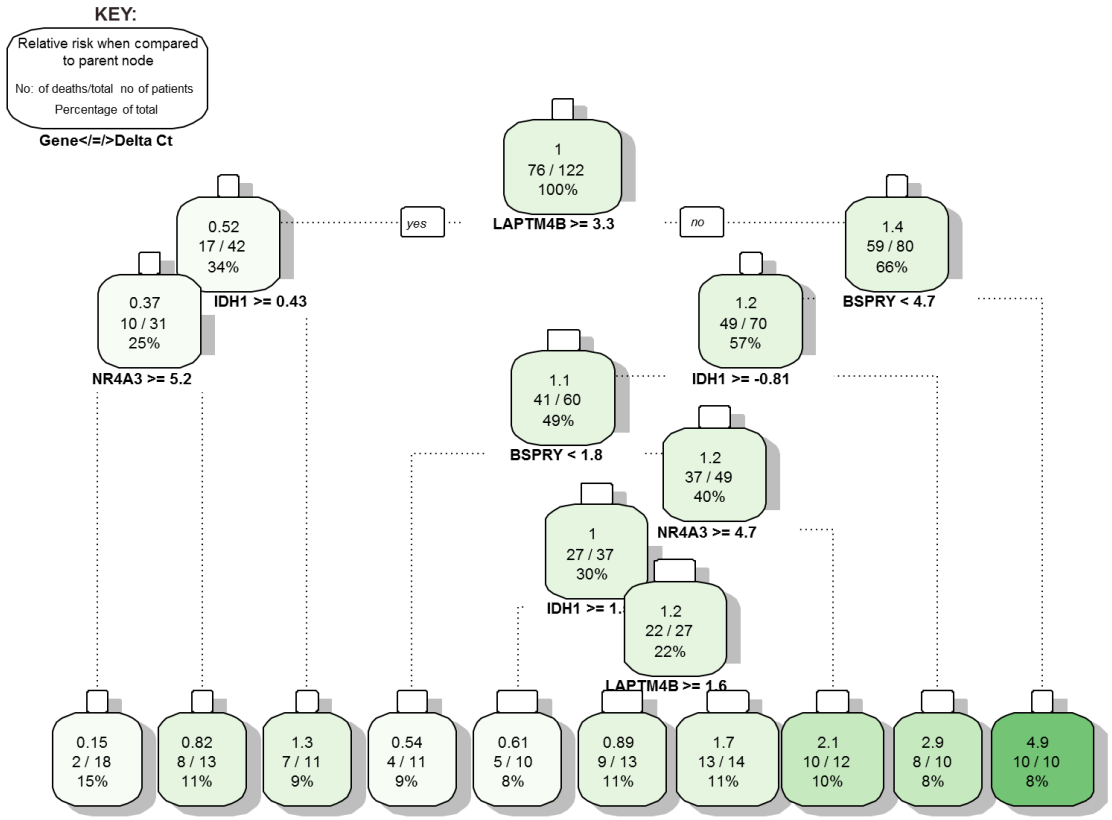


Figure 4.6 Preliminary classification tree for Overall survival in AML. The top five important genes across all four survival outcomes were used to generate a classification tree for Overall survival. (**Gene cut-off expressed in Delta Ct,** Yes to the left and No to the right)

Kaplan Meier analysis on the four groups (terminal nodes) from the pruned tree showed that two groups (*LAPTM4B* < 3.3, *BSPRY* < 4.7, *IDH1* < -0.81 and *LAPTM4B* < 3.3, *BSPRY* > 4.7) had very similar outcomes prompting us to combine them into a single group (data not shown). This modification in classification strategy resulted in three risk groups defined as follows: Favourable risk (*LAPTM4B* > 3.3); Intermediate risk (*LAPTM4B* < 3.3, *BSPRY* <4.7, *IDH1* >-0.81) and Adverse Risk (*LAPTM4B* < 3.3, *BSPRY* < 4.7, *IDH1* < -0.81; or *LAPTM4B* < 3.3, *BSPRY* > 4.7). This tree will be referred to as the *LAPTM4B/BSPRY/IDH1* (LBI) classification tree from here on (Figure 4.7).

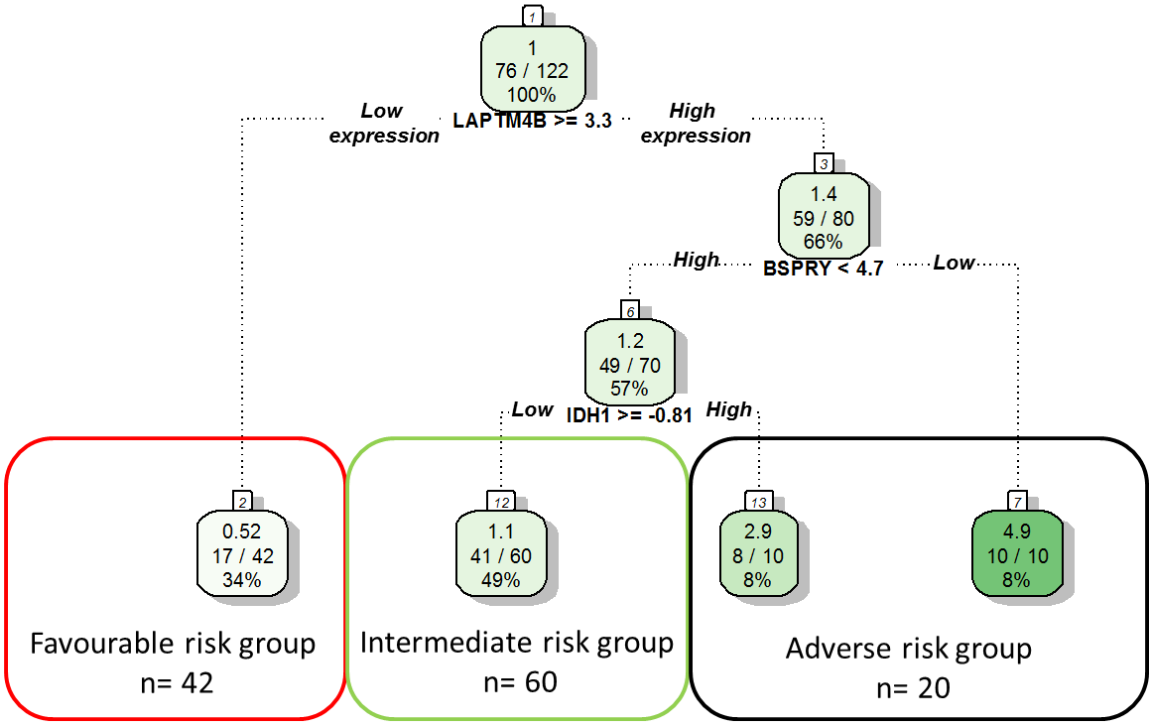


Figure 4.7 Pruned rpart classification tree segregates patients into three risk groups. The pruned tree avoids overfitting and utilises expression of *LAPTM4B*, *BSPRY* and *IDH1* to predict risk. (**Gene cut-off expressed in Delta Ct,** Yes to the left and No to the right)

4.2.4. Kaplan Meier analyses on the LBI Classification Tree

Kaplan Meier curves were plotted for the three AML groups as per the LBI classification system. The favourable group experienced significantly better overall survival ($p<0.0001$), disease specific survival ($p<0.0001$), disease free survival ($p<0.0001$) and event free survival ($p<0.0001$) than the medium risk groups and the latter experienced longer survival than the high risk group with overall significance for all four outcomes (Figure 4.10).

Kaplan Meier analysis was carried out on *LAPTM4B*, *BSPRY* and *IDH1* individually to assess the effect of each gene on survival. High and low groups were separated based on cut-offs determined by the LBI tree. For example, for *LAPTM4B*, low expression group consisted of 34% of patients as defined by a cut-off of $\Delta Ct \geq 3.3$. *LAPTM4B* expression was found to separate patients with better prognosis (Figure 4.8). Overall survival was different between *LAPTM4B* low and high groups with median survival of 196 and 1318 days, respectively (overall log-rank $p=0.0003$; hazard ratio – 2.5). DFS was also found to be different with log-rank p value of <0.0001 (hazard ratio – 2.5).

Though *BSPRY* expression could predict OS and DFS in a statistically significant manner (p=0.001, hazard ratio – 0.37; and p=0.0069, hazard ratio – 0.44, respectively), a good separation of patients in subgroups with different prognosis for survivals was not seen. *IDH1* high and low expression groups had significantly different OS and DFS outcomes (p=0.0003, hazard ratio – 3.2; and 0.0012, hazard ratio – 3.1, respectively). When analysing median survival times of high and low groups, it appears that *LAPTM4B* is useful in separating patients with favourable prognosis whereas *IDH1* expression is useful at separating those with poor outcomes.

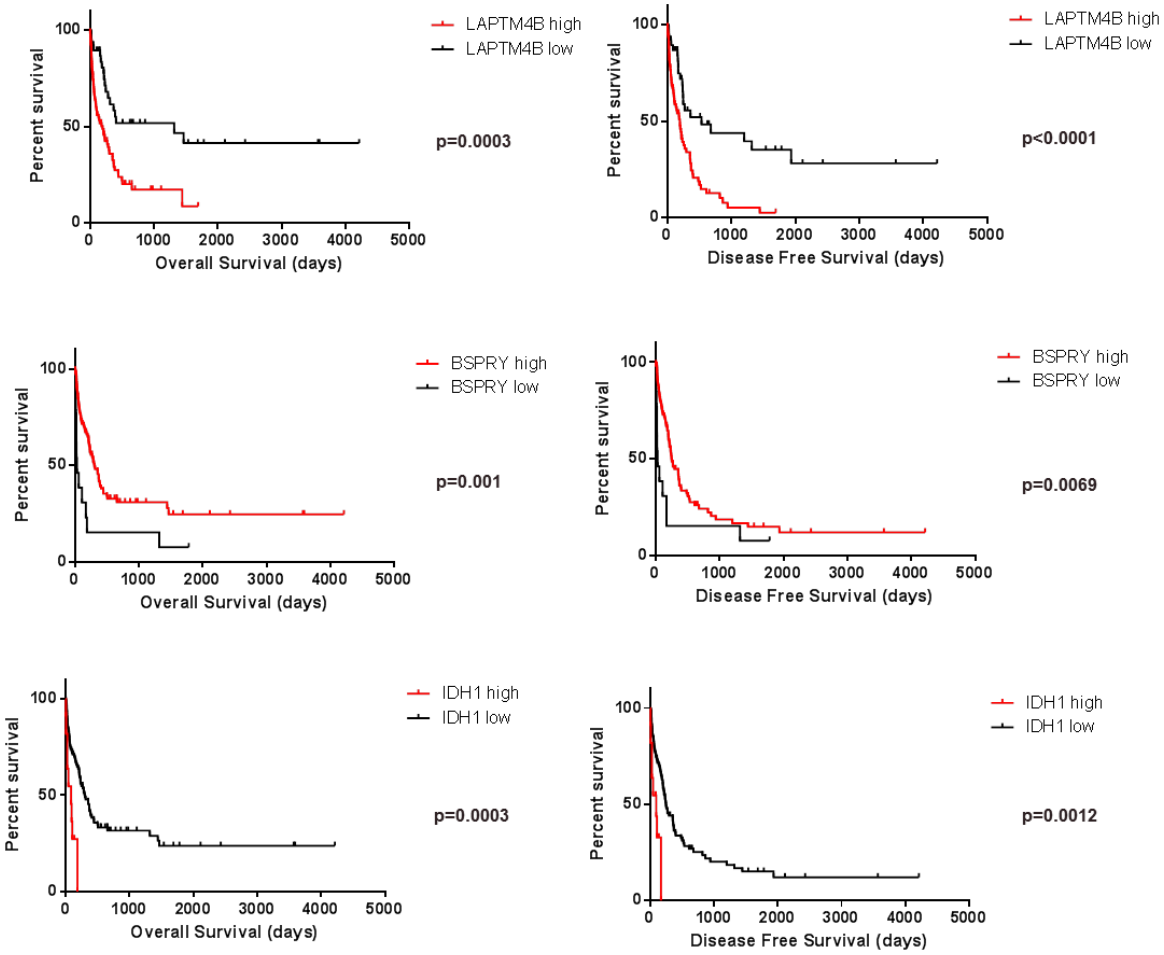


Figure 4.8 *LAPTM4B, BSPRY* and *IDH1* are predictors of OS and DFS. Log-rank test show significant differences in survival between high and low expression groups of *LAPTM4B, BSPRY* and *IDH1* for both OS and DFS outcomes.

4.2.5. Comparison with ELN Stratification Scheme

Having established that our decision tree could predict survival outcomes in our dataset efficiently, we set out to compare it with the existing ELN grouping. When applying the ELN stratification to our data, we saw good separation of the favourable group from the

intermediate-1, intermediate-2 and adverse groups which had very similar median survival in all four outcomes (Figure 4.10).

4.2.6. Development and Validation of a combined ELN/LBI Classification Scheme

To integrate our LBI decision tree into the ELN stratification scheme, we incorporated the ELN scheme as the first node in the decision tree and stratified the remaining ELN risk groups by the LBI classification tree (Figure 4.9). This combined ELN/LBI classification tree produced four risk groups. When assessed by Kaplan Meier analysis, importantly this new risk grouping resulted in excellent separation of all four risk categories with highly significant log-rank p values <0.0001 across all four outcomes (Figure 4.10). Median overall survival for favourable, intermediate-1, intermediate-2 and adverse groups were 2825, 407, 268 and 25 days, respectively. Combining the ELN risk model with our LBI model not only improved separation between risk groups, but also identified a group with extremely poor prognosis (black line). Of the 14 patients in this adverse group, only one showed even a partial response to induction therapy.

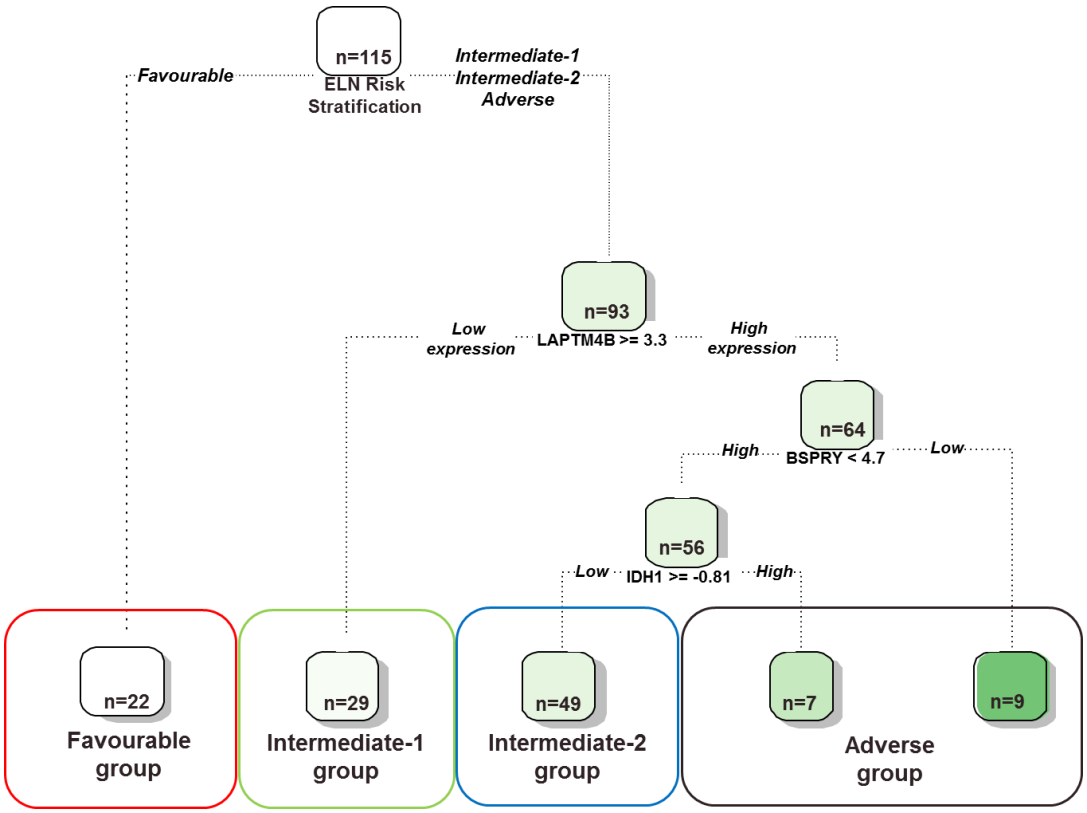


Figure 4.9 Integrated Risk Classification Strategy. Classification scheme developed by integrating the random forest decision tree into the ELN risk stratification scheme. (**Gene cut-off expressed in Delta Ct,** Yes to the left and No to the right)

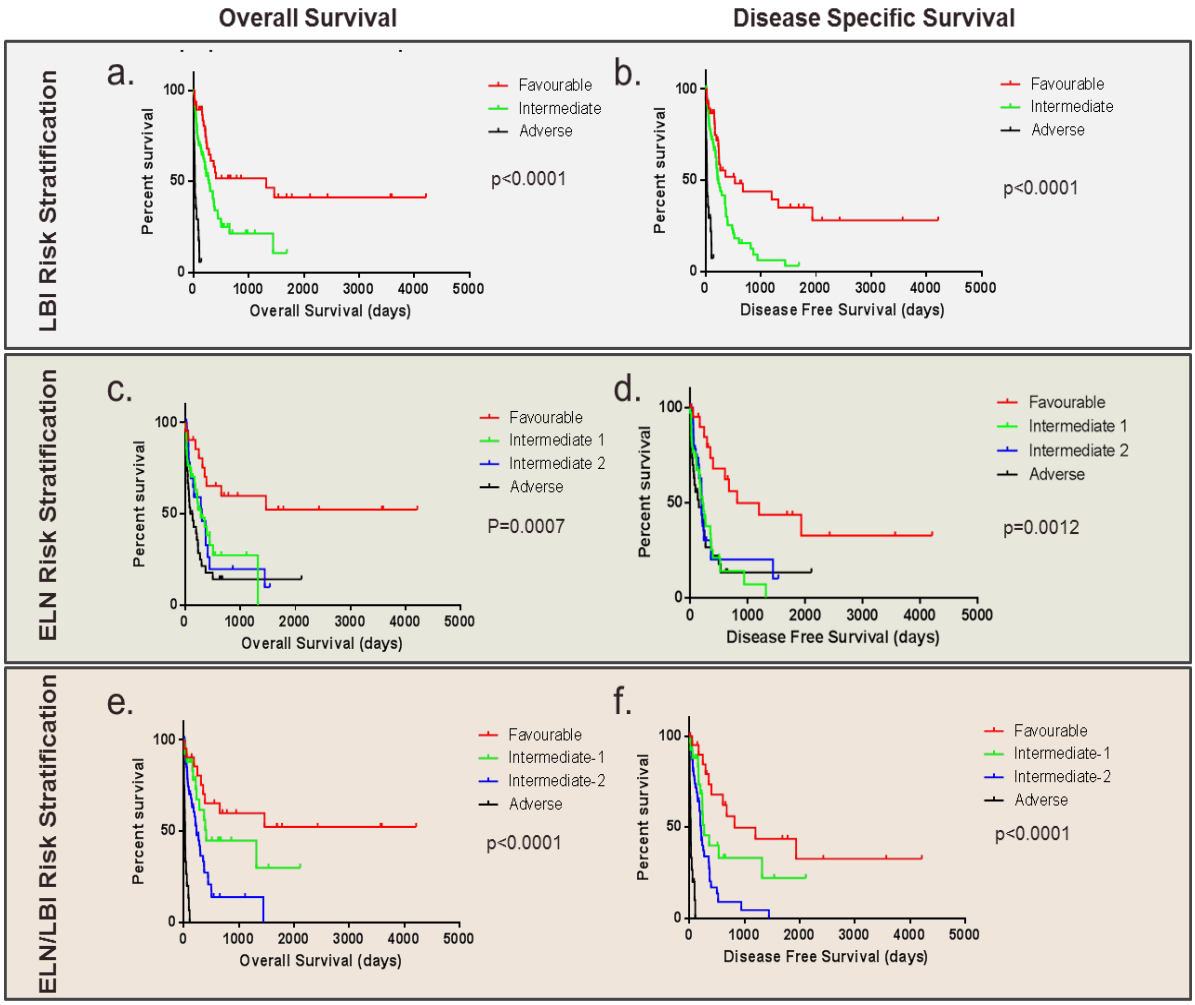


Figure 4.10 OS and DFS for LBI risk groups, ELN risk groups and the integrated ELN/LBI model. The integrated risk stratification strategy has a higher overall significance and better separation than ELN grouping. OS and DFS for LBI risk stratification (a and b, respectively), OS and DFS for ELN risk stratification (c and d, respectively) and OS and DFS for ELN/LBI model (e and f, respectively).

4.3. Discussion

We propose a risk classification strategy which can be integrated into the widely accepted ELN classification system that can substantially improve survival outcome prediction in AML patients. This integrated scheme has identified a group of patients with extremely poor prognosis in our cohort which was not picked up by ELN grouping. This is especially important in a clinical setting as these patients require novel/aggressive treatment strategies. Random forest analysis conducts internal cross-validation for each tree that is generated, wherein each tree is generated based on a random subset of the data (usually two thirds of total number of samples) and cross validates each tree using the remaining samples. It is still necessary to validate the integrated risk stratification strategy on a larger cohort of patients.

Random forest analysis is extremely useful in determining and visualising combinations of features (both categorical and continuous) which predict outcomes. It is especially useful when features (gene expression in this case) are continuous as it determines cut-offs based on outcome prediction rather than arbitrary cut-offs determined by the investigator. Random forest analysis incorporates randomisation at several levels and hence avoids bias introduced by other methods. Moreover, pruning of the preliminary decision tree helps avoid overfitting of data which is likely to happen with larger trees and also reduces the complexity of decisions.

Using random forest analyses, we have produced a classification tree model which can be combined with the ELN risk classification scheme to improve risk stratification. The ELN/LBI risk classification strategy appears similar to the 24 gene signature published by Li *et al.* in predicting survival, however, in order to assess this, it would be necessary to analyse our data using their strategy or vice versa(See Section 4.1) (103). Nevertheless, judging by the outcome of our current analysis, in which we identified a previously unidentified subset of patients with extremely poor outcome effectively, it appears that we can add significant value to the widely accepted ELN scheme by utilising expression from just three genes rather than 24 genes.

Interestingly, *LAPTM4B, BSPRY, ETFB, DAPK1*and *HEATR6* which have been highly ranked by random forest analysis for survival outcomes are also part of the 24 gene expression signature previously reported to be of prognostic significance (103). This 24 gene signature was derived from microarray data which is not as sensitive or quantitative as our qRT-PCR approach. The fact that *LAPTM4B* and *BSPRY* feature in the final tree, suggest that these genes may be the most important genes in the 24 gene signature that determine outcome and are probably sufficient to predict outcome. Our cohort not only acts as a validation cohort in confirming the importance of these genes in predicting outcome but also identifies the most important genes.

ERG expression levels are associated with outcome in both older patients and MLL-rearranged childhood AML (127, 128). *ERG* is among the top ten genes in all four outcomes and hence our findings confirm that *ERG* is important in determining survival outcomes. Similarly, abnormal *FLI1* expression is linked to adverse prognosis in AML (129). *FLI1* is among the top 25 genes in three survival outcomes in our random forest analysis.

MYC expression is reportedly a strong predictor of survival in t-AML (therapy related AML) and AML with MDS related changes (130). Prognostic significance of *MYC* is

likely to be limited to these groups as *MYC* ranked pretty low in our gene panel for all outcomes. Surprisingly, *GATA2* which has been previously reported as independent predictor of poor prognosis in AML was one of the lowest ranked genes in all four survival outcomes. Our gene panel consists of genes that are known or predicted to have an important role in myeloid disease biology and this might reflect on the other genes being of higher predictive value than *GATA2* (58).

Among the three genes used in the LBI tree, that together act as a powerful prediction tool, *LAPTM4B* was ranked as the most important. *LAPTM4B* resides on chromosome 8q22.1 and encodes lysosome-associated protein transmembrane-4β (LAPTM4B). Amplification of *LAPTM4B* has been associated with resistance to chemotherapy and recurrence in breast cancer (131). Trisomy 8 is a common chromosomal defect seen in AML and it has been proposed that amplification of *LAPTM4B* maybe one of the mechanisms contributing to AML pathogenesis. Polymorphisms in the 5' UTR of *LAPTM4B* have been implicated in increased susceptibility of breast cancer, malignant melanoma and non-small cell lung cancer (132-134). High *LAPTM4B* is also associated with subcellular localisation of cytotoxic drugs like anthracyclines, and it has been suggested that cytosolic retention of these drugs prevents them from damaging DNA (135). A recent study demonstrated that *LAPTM4B* expression can regulate intracellular localisation of ceramide. They demonstrated that high *LAPTM4B* expression resulted in clearance of ceramide from late endosomal organelles thereby stabilising late endosomal membranes and desensitising cells to lysosome-mediated death. This is also accompanied by an increased susceptibility to ceramide-dependent apoptosis. Several chemotherapeutic agents like tamoxifen and fenretinide that induce cell death in a ceramide-dependent manner are used to treat several solid cancers. AML $CD34^+$ and leukaemia stem cell-enriched AML $CD34^+CD38^-$ cells have been shown to be preferentially sensitive to fenretinide-induced apoptosis as opposed to normal $CD34^+$ cells (136). Hence, it may be beneficial to include drugs like fenretinide in the treatment regimen for *LAPTM4B* high AML patients.

BSPRY (B-box and SPRY-domain containing protein) is involved in controlling Ca^{2+} influx via negative modulation of TRPV5. AML cells are particularly sensitive to cellular Ca^{2+} levels and presumably Ca^{2+} modulation is one of the modes through which *BSPRY* expression contributes to disease and therefore survival of patients. There are Ca^{2+} targeted drugs in clinical trials which may prove useful upon a better understanding of the role of this pathway in disease biology (137, 138).

IDH1 (Isocitrate dehydrogenase 1) is a homodimeric enzyme that localises in the cytoplasm and peroxisomes. It catalyses the conversion of isocitrate to α-ketoglutarate in an $NADP^-$ dependent manner. *IDH1* mutations are recurrent events in both gliomas and AML. Pathogenesis of *IDH1/2* mutations is linked to altered substrate specificity such that the enzyme binds and converts α-ketoglutarate to 2-hydroxyglutarate in the presence of NADPH. Overexpression of wildtype IDH2 (mitochondrial homolog of IDH1) expression leads to higher levels of 2-hydoxyglutarate but it is unclear if increased expression of *IDH1* has a similar effect (108). In our data, high *IDH1* is a strong predictor of patients with extremely poor outcomes and this is consistent with other studies reporting *IDH1* to be an independent predictor of poor prognosis in AML (139).

Although individually, each of the three genes in the LBI classification tree are predictors of survival, when combined they form a powerful tool for identifying different risk groups for AML patients, rivalling current strategies. *LAPTM4B* is useful at identifying patients with favourable outcomes whereas *IDH1* seems to be useful in identifying a patient subset with extremely poor outcomes. *BSPRY* when used in combination with *LAPTM4B* and *IDH1* adds to their predictive power and helps in identification of the intermediate groups.

As per current practice, haematologists across the world use the ELN risk grouping or very similar strategies that utilise cytogenetic and molecular mutation detection assays to classify patients into prognostic groups at the time of diagnosis. This information along with other factors such as age and clinical data heavily influence treatment decisions. Inclusion of gene expression assays like qRT-PCR to assess the expression of the three genes (*LAPTM4B, BSPRY* and *IDH1*) with appropriate controls, can be implemented and can be used to further clarify prognostic risk stratification strategies that are currently in place. Though gene expression based assays are currently being used in the clinic to detect minimal residual disease in ALL and CML patients, it is

important to consider the ability to undertake well controlled and normalised RNA based assays in a standard diagnostic laboratory.

Validation of the findings from this study on completely independent cohorts is necessary before implementation in a clinical setting. This can be performed on publicly available microarray datasets available in databases like the Leukemia Gene Atlas (140). Though it is common practice to study gene expression profiles on unsorted peripheral blood or bone marrow (141), with simplification of cell purification techniques, it might be beneficial in the future to perform these assays on purified blasts (*e.g.* CD34+ or some other stem cell or leukaemic stem cell marker).

Chapter 5. Investigation of potential synthetic lethal interactions between oncogenic RAS and GATA2 in haematopoietic malignancy

5.1 Introduction

Ras proteins are small membrane associated GTP – binding proteins that act as molecular switches to control cellular processes like proliferation, survival and apoptosis when activated by extracellular signals. There are three known RAS isoforms – HRAS, KRAS and NRAS. Raf/MEK/ERK, PI3/Akt, RalGDS pathways are some of the downstream effectors of Ras signalling (Figure 5.1) (142). Functional activation of Ras mediated pathways is a common event in leukaemia. This can either happen via activating mutations in any of the Ras proteins or other proteins in the pathway or due to inactivation of negative regulators.

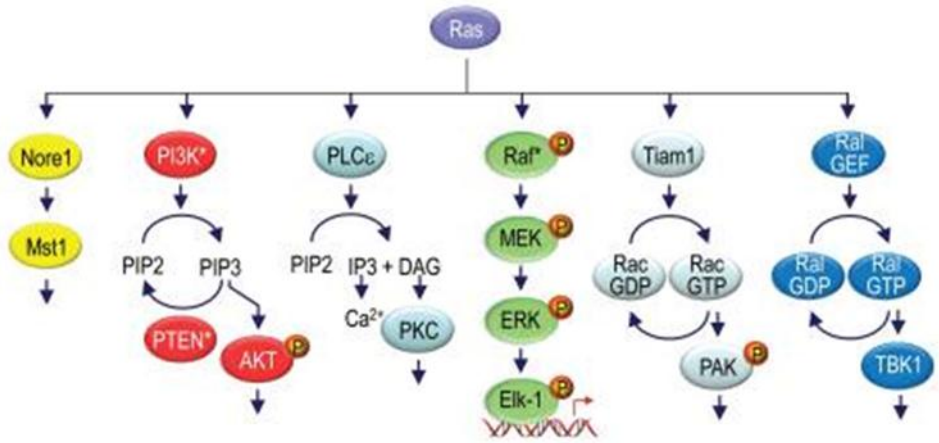


Figure 5.1 Ras mediated pathways implicated in oncogenesis. (Adapted from Cox *et al*, 2010) (143)

The leukaemogenic potential of oncogenic RAS has been studied in animal models previously utilizing bone marrow transplantation of transduced marrow as well as transgenic mice. Mutant Ras proteins have reduced GTPase activity and hence accumulate in GTP – bound confirmation resulting in constitutive activation. *NRAS* and *KRAS* mutations occur frequently in myeloid malignancies whereas *HRAS* mutations are rare. Expression of oncogenic *Nras* has been shown to cause AML or CMML depending on the viral titre used for transduction whereas *Kras* has been shown to cause only CMML (144). Expression of oncogenic RAS at endogenous levels leads to a block in terminal erythroid differentiation causing a mild increase in erythroid progenitors *in vitro* (145).

Recently, there have been reports on requirement of GATA2 activity in *KRAS* mutant colorectal cancer and non-small cell lung cancer. It has been demonstrated that loss of *GATA2* dramatically reduced tumour development in RAS mutant lung cancers (146, 147). We investigated any potential correlations between oncogenic *NRAS/KRAS* mutations and *GATA2* expression in our AML cohort (studied extensively in Chapter 4) and found interesting correlations (discussed in Section 5.2.1) which led us to study this interaction further. Hence, we decided to utilise a well-established mouse model of RAS induced disease in combination with conditional a *Gata2* knockout system to interrogate this interaction and understand whether they operate together in leukaemogenesis.

5.2 Results

5.2.1 Selective association of *RAS* mutations with *GATA2* low AML

We analysed *GATA2* expression in 156 AML samples (see Chapter 4 for details on samples and normalisation) by qRT-PCR, of which the RAS mutation status (*NRAS* and *KRAS*) was known for 122 samples. 8 CD34+ cells and 13 BMMNC from normal donors were also included. The AML samples were stratified into three groups by defining high and low groups using a cut-off of 95% confidence interval of the mean of *GATA2* expression in normal CD34+ cells. There were 15 AMLs with mutations in RAS – *NRAS* (11) and *KRAS* (5).

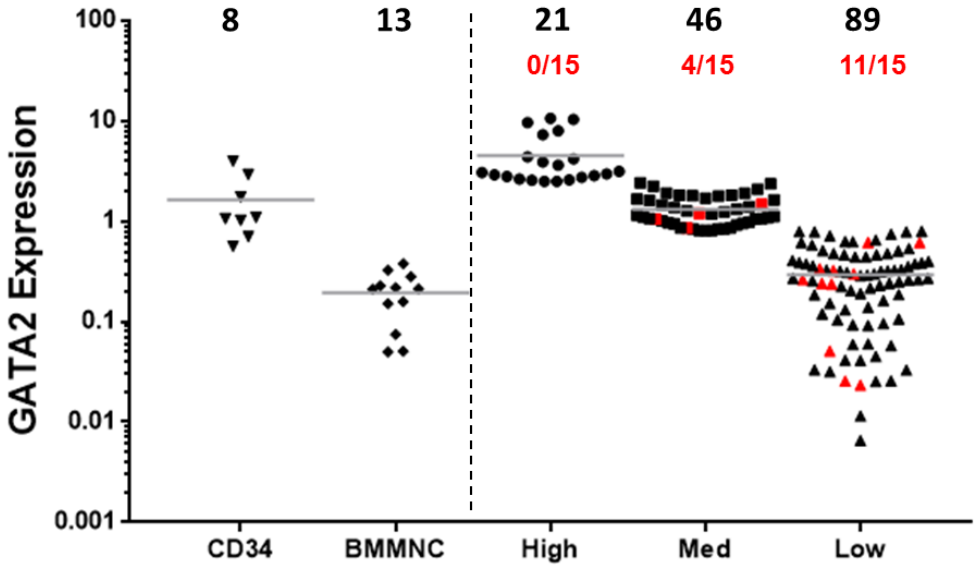


Figure 5.2 AML samples stratified on *GATA2* expression levels. *GATA2* mRNA expression levels were determined by qRT-PCR and samples were stratified as high, medium and low groups using a 95% confidence interval cut-off. *RAS* mutations (red triangles) are selectively represented in the low and medium *GATA2* groups. Number of *RAS* mutants of a total of 15 is shown in red. Total number of samples assessed for *GATA2* expression in each group written in black.

Interestingly 11/15 (73%) *RAS* mutant samples fell in the low *GATA2* group and while no *RAS* mutations were observed in the high *GATA2* group (Figure5.2). This suggests that *RAS* mutations in AML are predominantly found in an environment of low *GATA2* expression or that high *GATA2* expression possibly compensates for oncogenic RAS. Recent studies on *RAS* mutant lung and colon cancer cell lines have reported an absolute requirement for GATA2 activity and showed that knockdown of GATA2 suppressed growth of RAS mutant but not RAS wildtype cells (146, 147). Our data show residual GATA2 expression suggesting that AML cells must maintain at least

some GATA2 activity for oncogenicity of RAS. This raises the possibility, as shown in lung and colon cancer, that GATA2 is also a synthetic lethal partner for oncogenic RAS in AML.

5.2.2 NRAS G12D mice develop haematological disease

A pilot study was performed wherein we transduced bone marrow cells from donor mice with oncogenic *NRAS* G12D and transplanted them into lethally irradiated mice. Similar experiments have been done by several others previously with various outcomes (discussed briefly in Section 5.3). Owing to the variability in disease phenotypes and time of onset seen previously, we decided to test the system in order to determine the ideal transplantation dosage and establish assays to characterise disease phenotype. *NRAS* is more frequently mutated in AML and the G12 residue is the most recurrently mutated residue. Therefore we decided to use NRAS G12D for our studies.

Bone marrow was extracted from femurs, tibiae and pelvic bones of a single donor mouse of *Mx1-Cre* $^{-/-}$ *GATA2* $^{flox/+}$ background yielding 1.3 x 10^8 cells. These cells were then subjected to magnetic lineage depletion giving 4.2 x 10^7 cells. This step was carried out to eliminate terminally differentiated cells belonging to various lineages present in blood and enrich for stem/progenitor cells. Cells were plated in the presence of cytokines and retrovirally transduced to express either *NRAS* G12D or EV. Transduction efficiency was 3.3-7.4% for *NRAS* G12D and 24.3% for EV as analysed by flow cytometry (data not shown). Recipient mice were lethally irradiated and allowed to recover overnight before being transplanted with 1.4 x 10^6 cells via tail vain injections. The mice were weighed twice daily for 21 days and twice weekly after that to monitor for symptoms of haematological disease and health.

All transplanted mice showed engraftment at 3 weeks post transplantation. Peripheral blood analysis showed GFP positivity ranging from 16-42% for *NRAS* G12D transplanted mice and 31-70% for control mice transplanted with EV cells (Figure 5.3). Blood parameters were within normal ranges for EV recipients whereas five out of six *NRAS* G12D recipients displayed erythropaenia and neutropaenia at 3 weeks (Figure 5.4). One remaining *NRAS* G12D mouse displayed marked neutrophilia with other parameters within normal ranges.

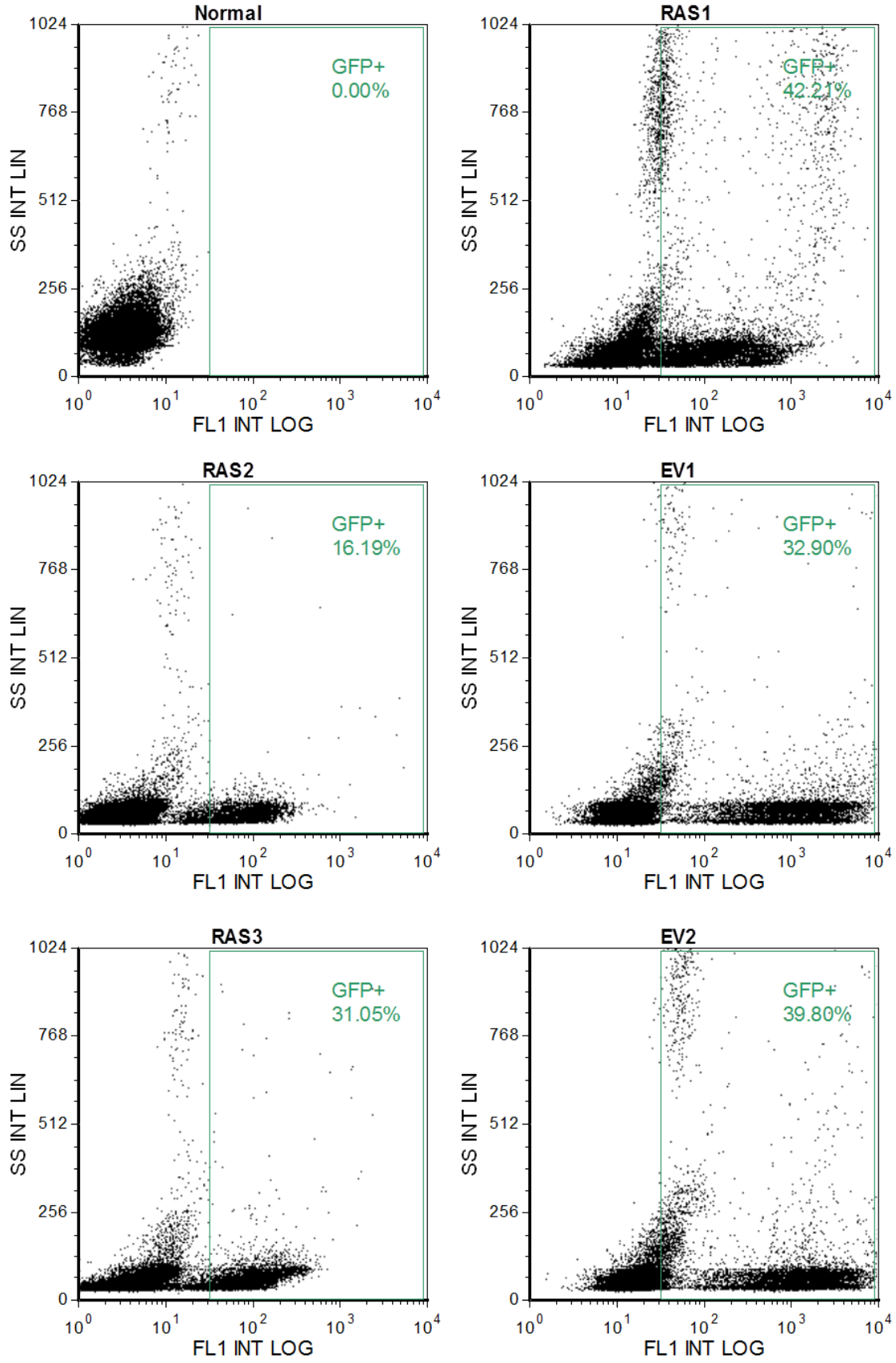


Figure 5.3 Successful engraftment in transplanted mice. GFP expression in peripheral blood of recipient mice in the pilot experiment, 3 weeks post transplantation show successful engraftment of transplanted cells.

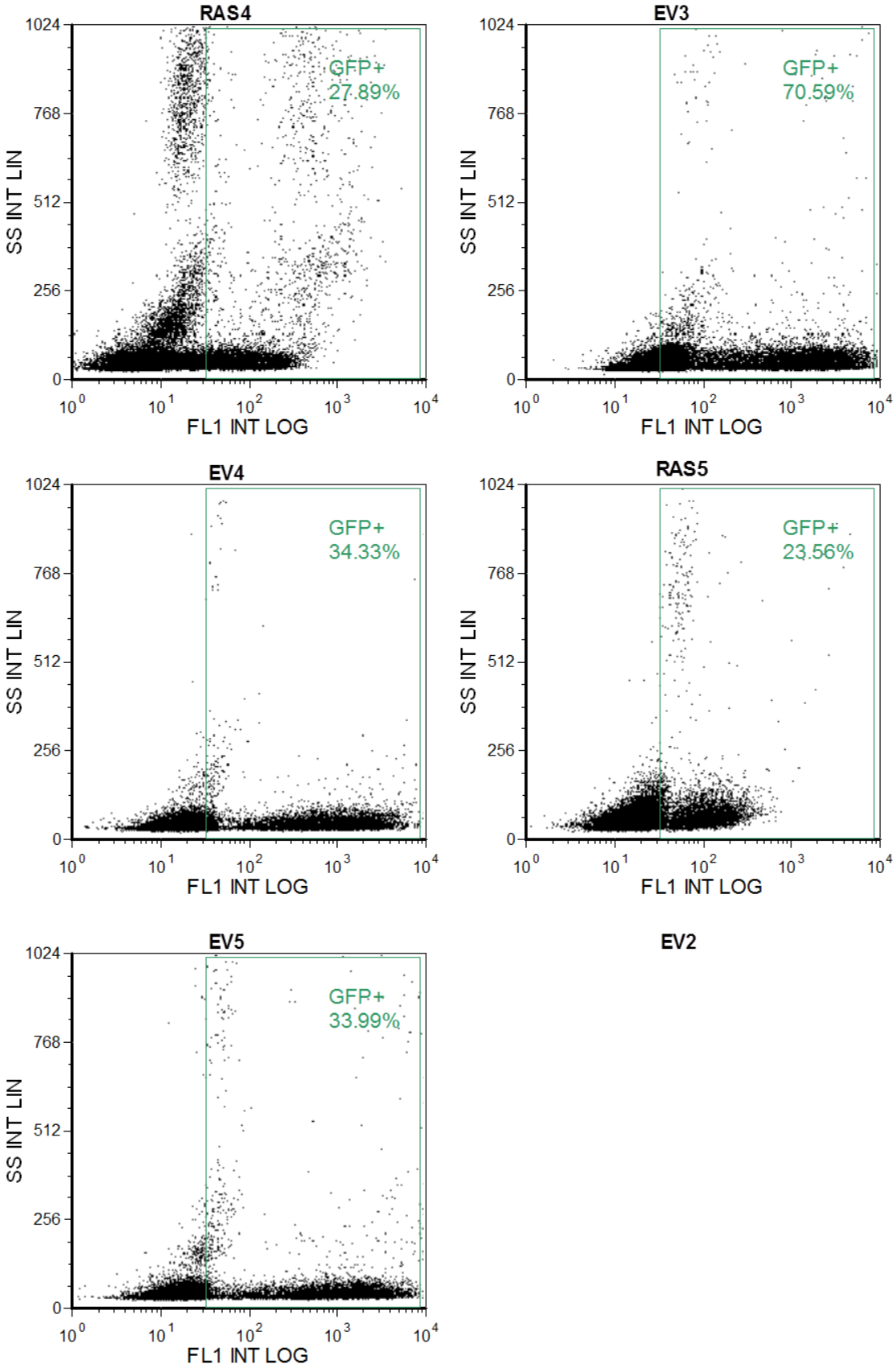


Figure 5.3 (continued)

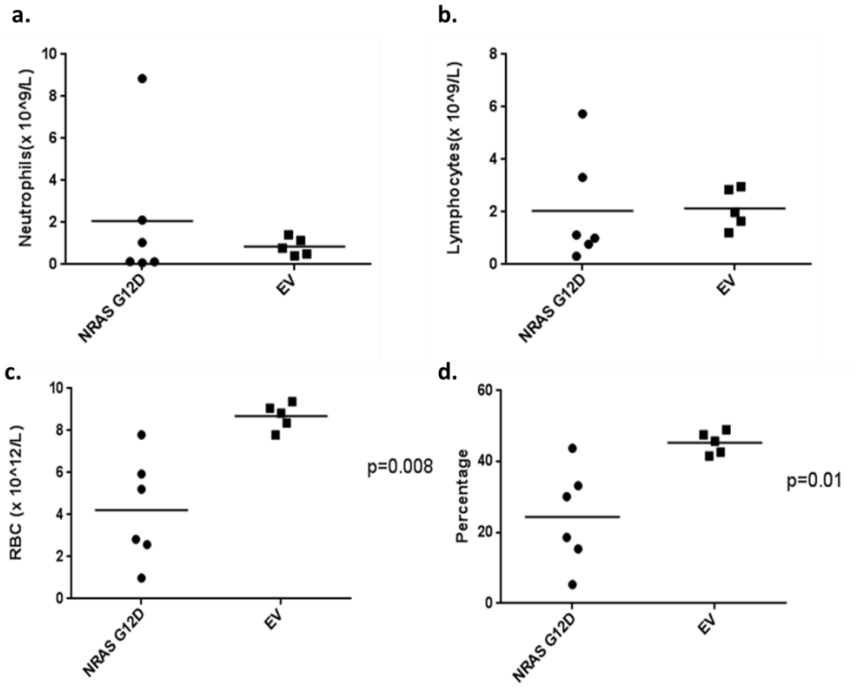


Figure 5.4 *NRAS* G12D recipient mice displayed significantly lower RBC counts and haematocrit than EV control mice. Haematological parameters of *NRAS* G12D recipient mice prior to being sacrificed in comparison with EV recipient mice at 3 weeks post transplantation. (a) Neutrophils, (b) lymphocytes, (c) erythrocytes and (d) haematocrit.

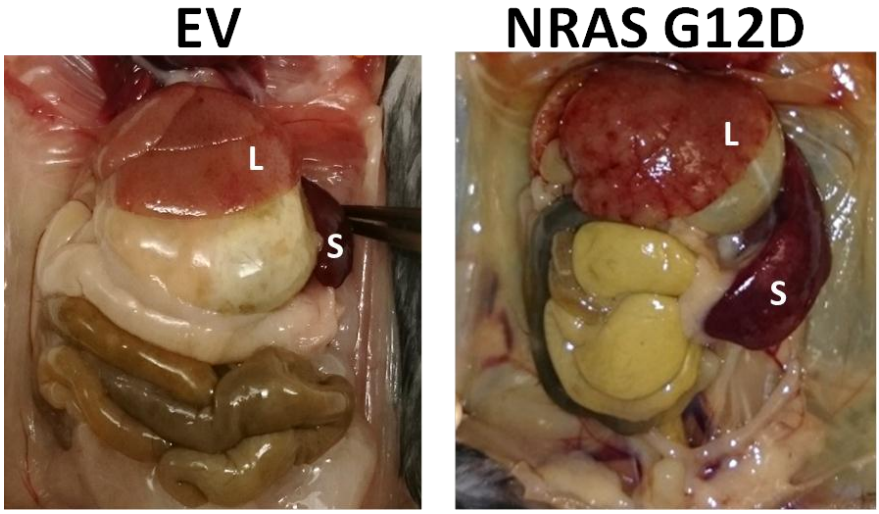


Figure 5.5 Abdominal/thoracic cavity of EV control mouse and *NRAS* G12D diseased mouse. Liver (L) appears abnormal and spleen (S) is markedly enlarged in the *NRAS* G12D mouse.

NRAS G12D recipients developed symptoms of haematopoietic disease by day 22-35. They all displayed hepatosplenomegaly (Figure 5.5) and their urine was deep yellow. Livers, apart from being slightly enlarged, had marked differences in gross structure. The spleens were 2-5 times the weight of normal spleens and had uneven contours (Table 5.1, Figure 5.6).

Table 5.1 Weights of spleens and livers from transplant recipient mice from pilot experiment

Mouse ID	Spleen weight	Liver weight
EV1	74 mg	0.9 g
EV2	81 mg	1.11 g
RAS1	525 mg	1.3 g
RAS2	407 mg	1.317 g
RAS3	425 mg	N/A
RAS4	198 mg	1.01 g
RAS5	307 mg	1.26 g
RAS6	350 mg	1.2 g

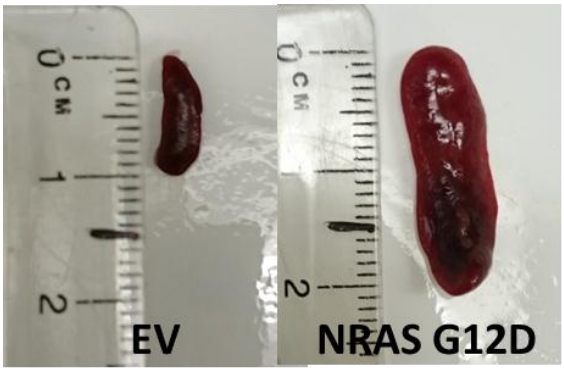


Figure 5.6 Diseased *NRAS* G12D mice exhibited marked splenomegaly. Spleens of *NRAS* G12D diseased mice were markedly enlarged and were at least 2-7 times the weight of an EV control spleen.

Haematoxylin and eosin (H&E) sections of the spleen revealed follicular enlargement and effacement of normal splenic architecture by proliferating neoplastic cells (Figure 5.7). The cells were a pleomorphic population of lymphoblast like cells with large round to ovoid nuclei while a few nuclei had clefts. Some cells had elongated nuclei resembling histiocytes. Multifocal necrosis was seen within foci possibly due to relatively high mitotic rate leading to outstripping of blood supply. There were also a few instances of abnormal mitosis consistent with what is commonly seen in neoplastic lesions.

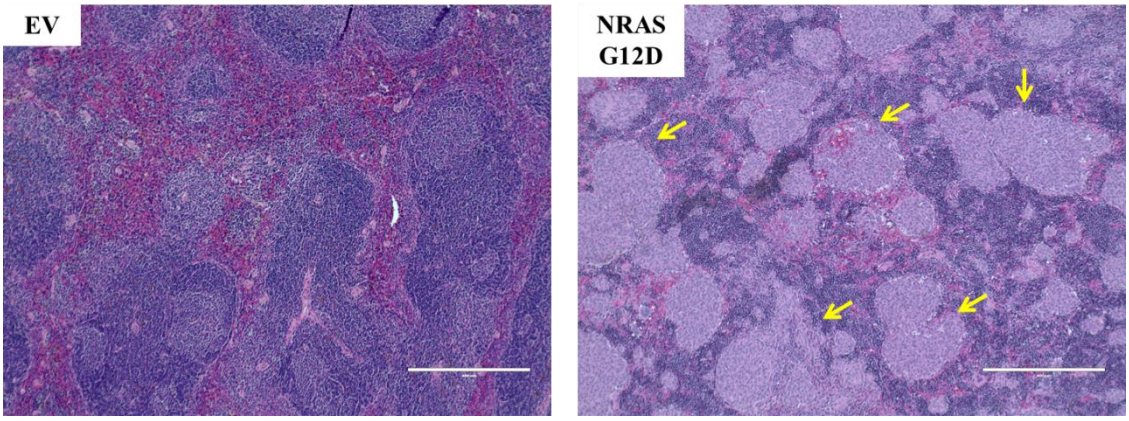


Figure 5.7 Spleen of *NRAS* G12D diseased mouse displays disrupted architecture and abnormal morphology. Follicular enlargement and effacement of normal splenic architecture (yellow arrows) in H&E stained sections of *NRAS* G12D diseased spleen. EV controls displayed normal architecture (Magnification 10X)

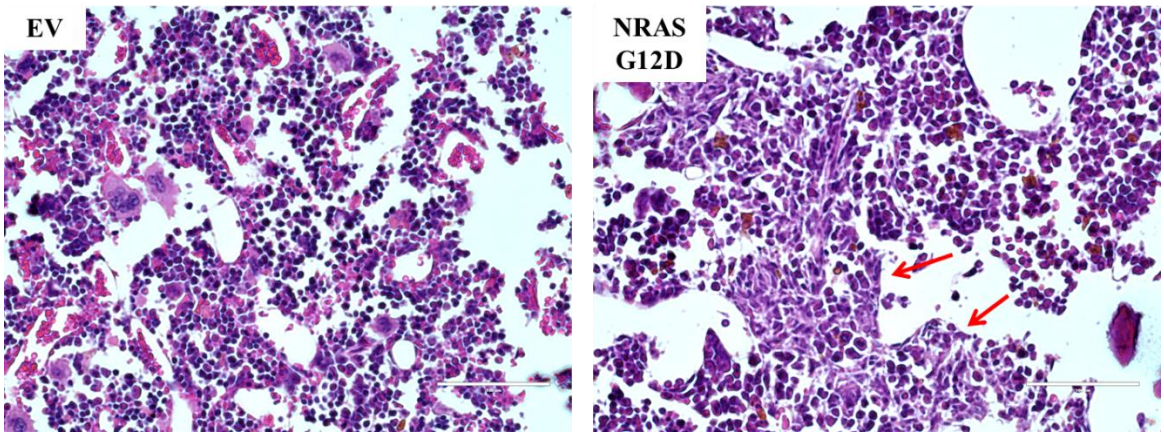


Figure 5.8 Diseased *NRAS* G12D marrow contained cells of abnormal morphology. H&E stained sections of the sternum showed the presence of abnormal lymphoblast like cells (red arrows) within marrow which were absent in marrow from EV transplanted mice. (Magnification 20X)

Stained sections of the sternum showed infiltrating cells colonising the marrow of *NRAS* G12D mice (Figure 5.8). Infiltrating cells within the marrow had a more histiocytic appearance when compared to morphology of cells infiltrating other tissues. Marrow and spleen cells of sick mice had >30% GFP positivity when analysed by flow cytometry.

The liver showed heavy periportal infiltration of lymphoblast like cells (Figure 5.9). Livers from EV control mice (n=2) showed evidence of mild steatosis which is likely to be a consequence of hepatic injury following lethal irradiation. *NRAS* G12D mice (n=6) did not show any evidence of steatosis though they were treated identically to EV control mice. The reason for this is unclear.

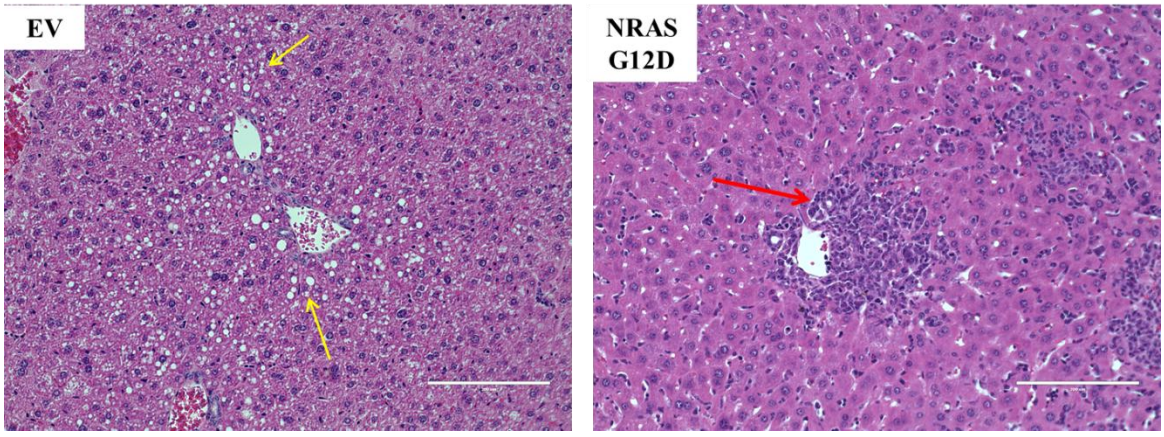


Figure 5.9 Liver of *NRAS* G12D diseased mice shows periportal infiltration. H&E stained section of *NRAS* G12D diseased liver displayed periportal infiltration (red arrows). EV transplanted liver shows evidence of mild steatosis (fatty liver) indicated by lipid droplets (yellow arrows). (Magnification 10X)

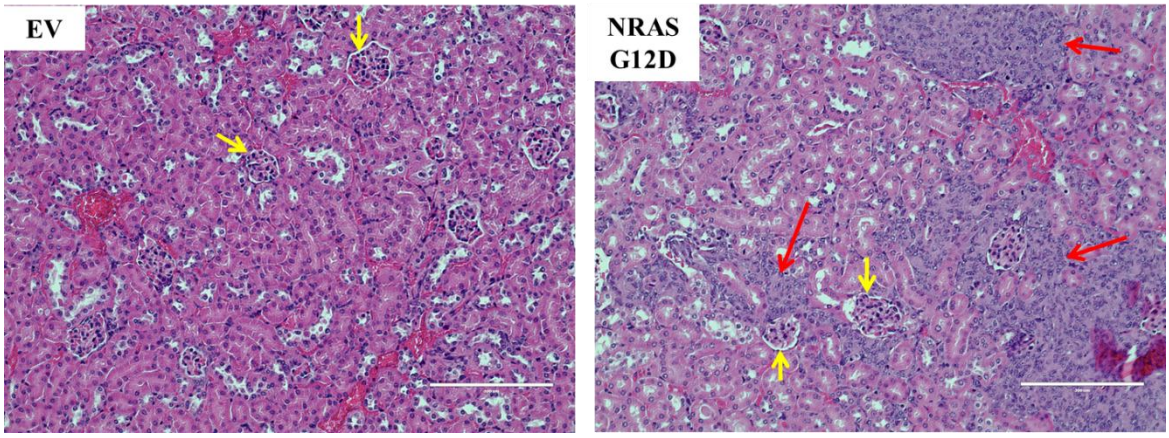


Figure 5.10 Extensive tumour infiltration into the renal cortex in *NRAS* G12D diseased mice. H&E stained section of *NRAS* G12D diseased kidney shows tumour foci (red arrows) in the cortical interstitium. Glomeruli are highlighted by yellow arrows. (Magnification 10X)

Kidneys from one of the recipients appeared abnormal on macroscopic examination with small white nodules on the surface. H&E sections confirmed heavy tumour infiltration especially into the cortical interstitial region (Figure 5.10).

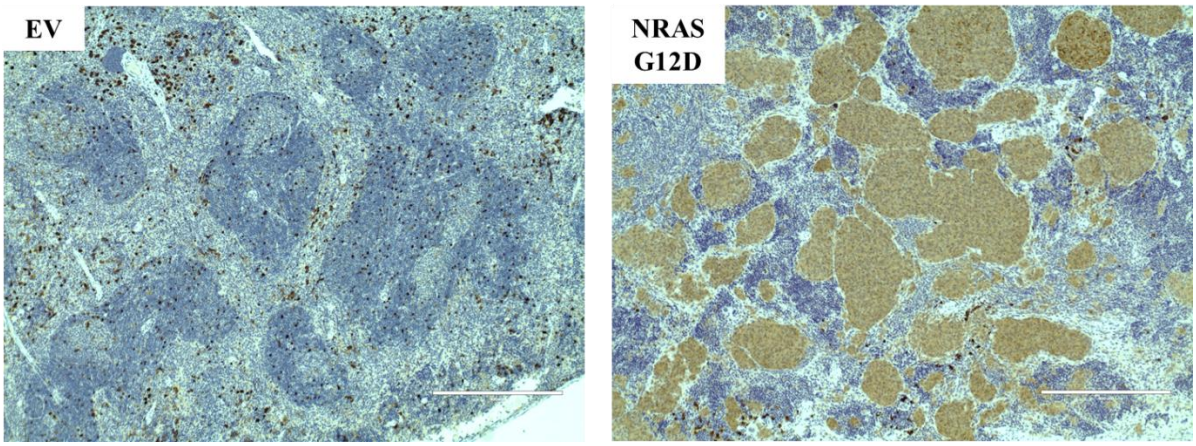


Figure 5.11 Tumours originate from transplanted GFP positive cells. Tumour foci in spleen stain positive in immunohistochemistry for GFP whereas EV control spleen has scattered GFP positivity indicative of normal tissue as expected. (Magnification 10X)

Morphological features were highly suggestive of B cell lymphoma or histiocytic sarcoma or possibly a lymphoma with histiocytic differentiation. Immunohistochemical staining for GFP confirmed that the tumours were formed by the transplanted GFP positive cells (Figure 5.11). Immunohistochemical staining with B220 and CD3 was carried out to confirm type of haematological malignancy and were both found to be negative (staining excluded tumour foci). Further staining for Mac-1, F4/80 and Mac-2 is currently underway to identify myeloid or histiocytic components within the foci.

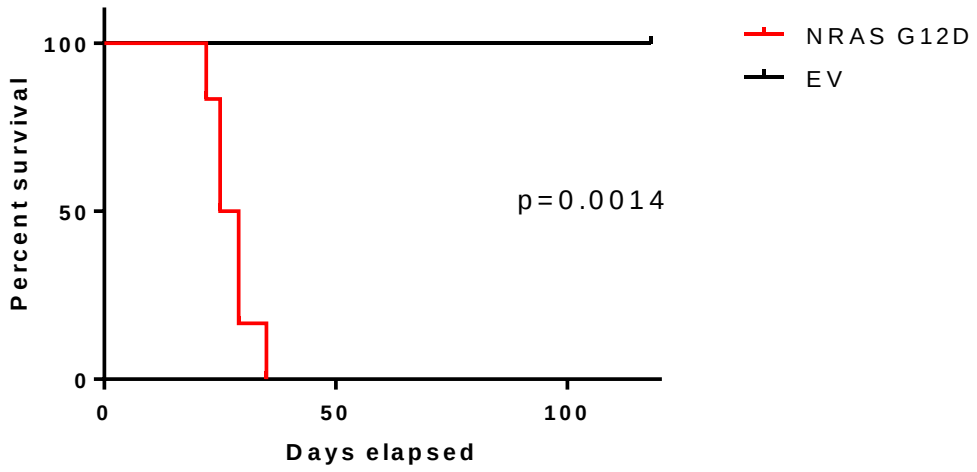


Figure 5.12 *NRAS* G12D recipient mice have significantly shorter survival times than EV recipients. Kaplan Meier curve showing survival of *NRAS* G12D recipient mice and EV mice.

Kaplan Meier analyses showed significant differences (logrank test, p=0.0014) in survival between *NRAS* G12D recipients and EV recipients (Figure 5.12). All *NRAS* G12D mice had to be humanely killed by day 35 as they developed signs of disease which included weight loss, hunched posture, ruffled coat and reluctance to move. All

EV recipients are healthy and show no signs of haematological disease at >120 days post transplantation.

5.2.3 *NRAS* G12D Transduced Lineage Depleted Bone Marrow Cells have *in vitro* Impaired Differentiation

NRAS G12D transduced lineage depleted bone marrow cells and EV controls were sorted on GFP positivity (data not shown) and plated at 2500 cells/ml MethoCult differentiation media. Colonies were scored on day 7. *NRAS* G12D cells showed increased number of BFU-E and CFU-GEMM colonies along with a significant decrease in CFU-G colonies when compared to EV cells (Figure 5.13). Individual colonies were scored and several colonies were stained with Giemsa to confirm colony types (data not shown).

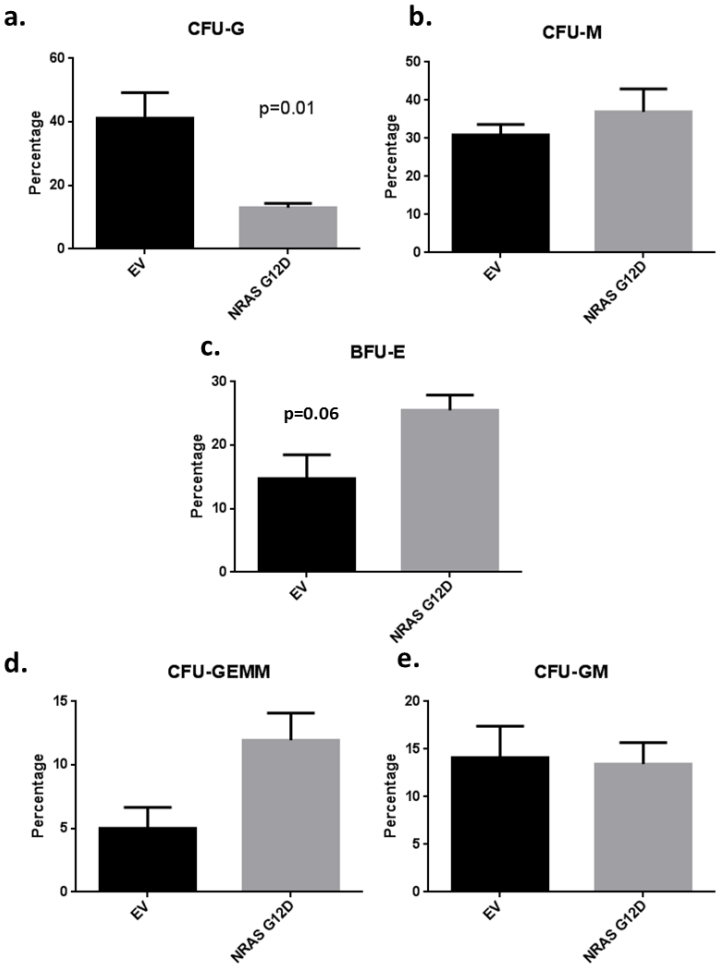


Figure 5.13 *NRAS* G12D impacts haematopoietic differentiation. The number of colonies formed per 2500 cells plated reported as percentage of total colony number. Statistical significance was determined by Mann Whitney test.

5.2.4 Do NRAS G12D and GATA2 cooperate in leukaemogenesis?

The pilot experiment served as proof of concept that we were able to induce haematological disease in recipient mice by transducing with *NRAS* G12D prior to transplantation. Our aim was to investigate the interdependence of oncogenic RAS and GATA2 levels in haematopoietic disease. Based on the outcome of the pilot experiment, wherein recipients of *NRAS* G12D transduced cells developed disease with extremely short latency, the number of *NRAS* G12D transduced cells transplanted into lethally irradiated recipient mice in this phase was reduced to 0.45 – 0.47%. This was done to slow the onset of disease considerably to enable us to observe acceleration of disease, if any, in the absence of GATA2. Donor mice were Mx1 Cre positive and either *Gata2*$^{+/+}$ (WT) or *Gata2*$^{Flox/+}$ (Het). Both donors yielded ~1 x 10^8 bone marrow cells and gave 1.3 x 10^7 and 1.9 x 10^7 cells after lineage depletion, respectively. A total of 1 x 10^6 cells were transplanted into each lethally irradiated recipient mouse. Mice were weighed twice daily for 21 days and twice weekly after that and monitored for general wellbeing. On day 21, mice were bled for GFP analysis and complete blood counts. Following that, mice were given three doses of pIpC with two day intervals to induce the excision of floxed *GATA2* (148).

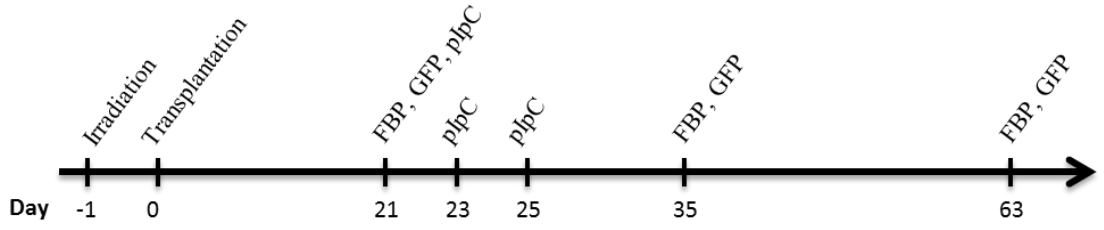


Figure 5.14 Timeline for NRAS G12D - GATA2 cooperation experiment

Mice were bled at 5 weeks, 9 weeks, 13 weeks post transplantation. GFP positivity in peripheral blood at various time-points is summarised in Table 1. Though haematological parameters consistently improved in all mice over the weeks following transplantation, there appeared to be a trend of decreasing GFP positivity in all the mice irrespective of donor background or construct (Figure 5.15). Counts at 3, 5 and 9 weeks after transplantation are plotted in Figure 5.16. The reduction in GFP positivity may reflect the inability to successfully transduce long term repopulating stem cells. Increasing cell numbers suggest that there definitely were stem/progenitor cells within the untransduced fraction which effectively repopulated the mice. Detailed blood counts at various time-points have been listed in Appendix VIII.

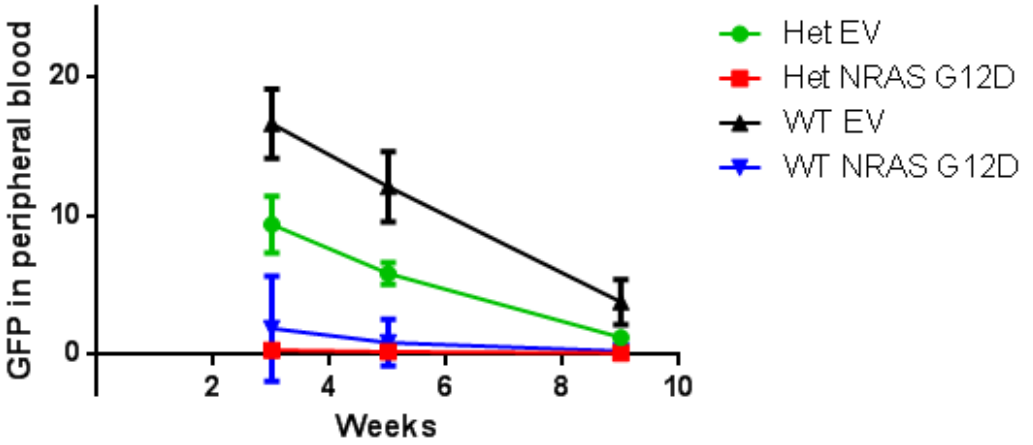


Figure 5.15 Percentage of GFP positive cells in peripheral blood decrease with time. Flow cytometric analyses of peripheral blood shows a progressive reduction in GFP positive cells in all mice irrespective of donor background or construct.

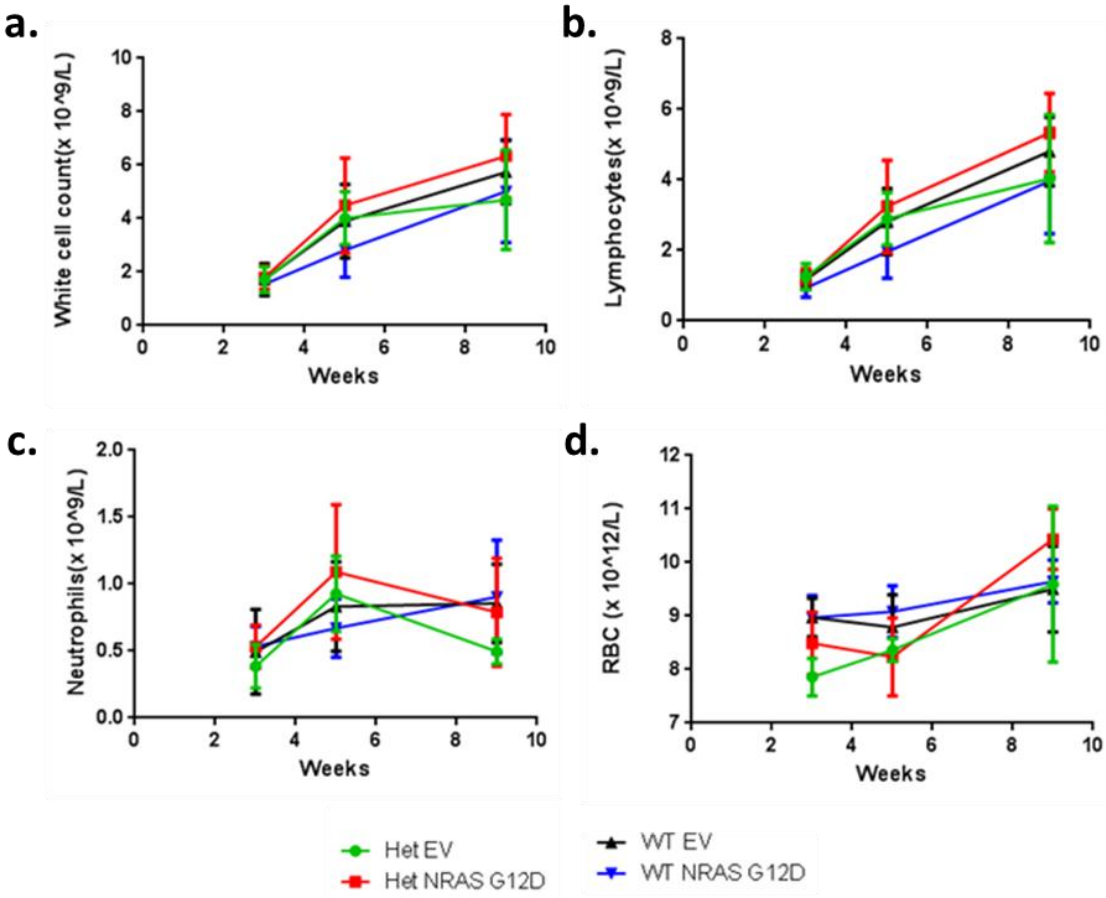


Figure 5.16 Haematological parameters rise to normal ranges in weeks following transplantation. Complete blood pictures show steady increase in blood counts attaining normalcy by 9 weeks. (a) White cell count, (b) Lymphocytes, (c) Neutrophils and (d) Erythrocytes.

5.3 Discussion

Studies on *GATA2* expression in AML have only identified *GATA2* high expressers as prognostically significant (58). This is because most studies use normal mononuclear cells (MNCs) as the reference to stratify AMLs. *GATA2* expression in MNCs is relatively low when compared to normal CD34+ cells and AML (Figure 5.2.1). We stratified AML samples into *GATA2* expression groups using normal CD34+ cells as the reference group. We argue that AML blasts are usually CD34+ and exhibit stem cell-like properties and hence normal CD34+ cells may serve as a better reference when evaluating AML samples. This approach helped identify a previously unrecognized group of low *GATA2* expressers that had an over-representation of *RAS* mutations.

NRAS G12D transduced cells showed a significant decrease in CFU-G colonies which is in contrast to overexpression of wildtype RAS that has been shown to cause altered lineage commitment from monocyte to granulocyte lineage in a mouse model (149). The decrease in CFU-G colonies also seems to be linked with a strong trend towards BFU-E and a small increase in CFU-GEMM. This is suggestive of altered lineage commitment and a differentiation block in the generation of CFU-G colonies from the CFU-GM stage (Figure 5.17).

Though there was a near-significant increase in BFU-E colonies in *in vitro* differentiation assays in the presence of *NRAS* G12D, all the *NRAS* G12D recipient mice in the pilot experiment had low red cell counts. This is consistent with a report from Li *et al*. where it was demonstrated that expression of *NRAS* G12D from the endogenous locus impairs erythroid differentiation (150). These findings together indicate that a differentiation block may also exist downstream of the erythroid progenitors thereby preventing terminal erythroid differentiation (Figure 5.16). *GATA2* is highly expressed in erythroid precursors during normal differentiation and hence it would be interesting to see how NRAS G12D mediated differentiation defects are affected by lowering GATA2 expression (151). One might postulate that lowering GATA2 levels in addition to expression of oncogenic *NRAS* G12D would further perturb erythroid differentiation.

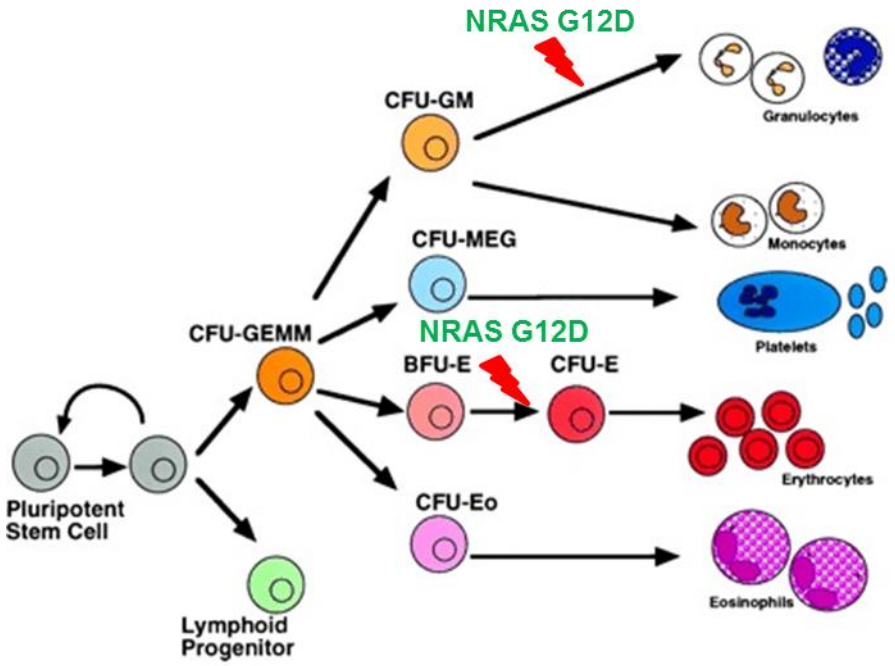


Figure 5.17 NRAS G12D alters lineage commitment in haematopoietic differentiation. (Adapted from Socolovsky, 1998) (152)

NRAS G12D overexpression models have been shown to cause CMML like disease and AML (153). While heterozygous expression of *NRAS* G12D from the endogenous locus can induce a spectrum of fatal haematopoietic disorders like myeloproliferative disease (MPD), MDS like disease, lymphoproliferative disease concomitant with myeloid disease and histiocytic sarcoma; homozygous expression of *NRAS* G12D in the endogenous locus causes T cell lymphoblastic leukaemia (TALL) in mice (150, 154). In our pilot experiment, wherein lethally irradiated mice were transplanted with marrow overexpressing *NRAS* G12D, mice rapidly developed haematological disease with lymphoblastic/histiocytic morphology.

The mice in the pilot experiment developed haematologic disease with extremely short latency periods (22-35 days) compared to previously reported (40-80 days) (153). Studying cooperation of GATA2 expression would not be possible if the mice developed *NRAS* G12D induced disease in such a short span. Hence, we adjusted conditions to achieve 10-20 times fewer *NRAS* G12D transduced-GFP positive cells compared to the pilot experiment which is in closer alignment to a previous study where disease onset was at ~6weeks (153). However, the transplanted mice showed a drop in GFP positivity over time irrespective of the construct. This suggests that very few long term repopulating stem cells has been transduced. This experiment will be repeated ensuring efficient transduction of the haematopoietic stem cell compartment.

Once established, this model will help us understand the effect of *Gata2* knockdown on the oncogenicity of mutant *NRAS*. Kumar *et al.* demonstrated that siRNA knockdown of *GATA2* in non-small cell lung cancer (NSCLC) cell lines carrying KRAS mutations reduced viability and increased apoptosis *in vitro*; and completely abrogated tumour formation in xenograft models (147). It seems counter-intuitive that in our AML cohort we see low *GATA2* expression in AMLs with RAS mutations but we predict that further reduction in *GATA2* levels would be lethal to the cells. Moreover, the absence of two *GATA2* mutations in the same AML clone is yet another indicator that haematopoietic cells require at least one functional copy of *GATA2* to survive. It would be interesting to see if lowering *Gata2* levels in *NRAS* G12D transduced cells accelerates or slows disease onset or leads to a different type of haematopoietic malignancy in our mouse model. If oncogenicity of NRAS is somehow dependent on *GATA2* expression, this could open avenues for new therapeutic interventions. Although GATA2 itself is not currently druggable, it may be possible to target GATA2 regulated pathways like the proteasome and Rho signalling, to exploit any such synthetic lethality as described earlier (147).

RAS mutations occur in 12-27% of patients with AML (155). A recent report which states that the majority of normal karyotype AMLs are low expressers of *GATA2* confirms our observation that 57% AML patients also exhibit low *GATA2* expression (156). These findings highlight that studying the role of *GATA2* expression and *RAS* mutation in AML may have implications on a large number AML patients. Furthermore, studying the underlying mechanisms by dissecting transcriptional activity and signalling pathways, in each scenario will improve our understanding of AML biology.

5.4 Recommendations for future work

We will continue with RAS/GATA2 bone marrow transplantation studies to identify rate of onset and phenotype of malignancy. It will be important to characterize the type of haematopoietic disease that the mice develop using immunohistochemistry for specific markers (as described earlier) and flow cytometry. We have established flow cytometry panels to assess peripheral blood (TER119, CD45, B220, Mac-1, Gr-1, CD3ε, CD5) and bone marrow (c-kit, Sca-1, CD150, CD48, CD41, Flt3, CD16/32, CD105, IL7Rα) of diseased mice to identify any resultant haematopoietic malignancy. It would be interesting to see how lowering *GATA2* levels in *NRAS* G12D transduced

cells affects the establishment and maintenance of haematopoietic malignancy. If knockdown of GATA2 proves to be effective in blunting the oncogenic effects of RAS, it would be worth investigating treatment options that either target GATA2 expression itself of pathways regulated by GATA2.

Chapter 6. Self-reverting Mutations Autocorrect the Clinical Phenotype in a Diamond Blackfan Anaemia Patient

6.1 Introduction:

We investigated the case of a young boy with a clinical diagnosis of Diamond Blackfan Anaemia with an unresolved genetic cause. We identified the causative genetic defect in the child using whole genome sequencing and high density SNP arrays. We also identified underlying somatic revertant mutations which we predicted would correct the DBA phenotype. This was consistent with the observation that the child attained spontaneous remission for a period of 18 months during which he was transfusion independent.

6.2 Statement of Authorship

Statement of Authorship

Title of Paper	Whole Genome Sequencing and High Density Arrays End the Diagnostic Odyssey for a Unique Diamond Blackfan Anaemia Patient
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Parvathy Venugopal		
Contribution to the Paper	Performed assays confirming genetic defect, manuscript preparation. Contributed to collation of clinical and genetic data, coordination of genetic testing, analysis and interpretation of WGS and MPS data for identification of deletion breakpoint.		
Overall percentage (%)	60%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis.		
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Sarah Moore		
Contribution to the Paper	SNP array analysis		
Signature		Date	26 FEB 2016

Name of Co-Author	David M Lawrence		
Contribution to the Paper	Copy number analysis for identification of deletion breakpoint		
Signature		Date	25/2/16

Name of Co-Author	Amee J George		
Contribution to the Paper	SBDS protein and RT-PCR experiments		
Signature		Date	23/2/2016

Name of Co-Author	Ross D Hannan		
Contribution to the Paper	Prior genetic testing and expert feedback on manuscript		
Signature		Date	29/2/2016

Name of Co-Author	Jinghua Feng		
Contribution to the Paper	Bioinformatics analysis of MPS data for identification of potential somatic variants		
Signature		Date	4/3/2016.

Name of Co-Author	Amanda Tirimacco		
Contribution to the Paper	TruSight Cancer panel		
Signature		Date	4/03/16

Name of Co-Author	Alexandra L Yeoman		
Contribution to the Paper	TruSight Myeloid panel		
Signature		Date	25/02/16.

Name of Co-Author	Chun Chun Young		
Contribution to the Paper	SNP array		
Signature		Date	25/2/16

Name of Co-Author	Andreas W Schreiber		
Contribution to the Paper	Bioinformatics analyses – pseudogene/gene mapping, identification of somatic variants		
Signature		Date	25/2/2016

Name of Co-Author	Christopher N Hahn		
Contribution to the Paper	Experimental design, data interpretation and manuscript preparation		
Signature		Date	25/02/2016

Name of Co-Author	Christopher Barnett		
Contribution to the Paper	Consulting clinician, sample collection, feedback on manuscript		
Signature		Date	23/2/16

Name of Co-Author	Ben Saxon		
Contribution to the Paper	Manuscript preparation, clinical care, consulting haematologist		
Signature		Date	23 FEB 16

Name of Co-Author	Hamish S Scott		
Contribution to the Paper	Experimental design, data interpretation and manuscript preparation		
Signature		Date	23/3/16

Name of Co-Author	Richard J. D'Andrea		
Contribution to the Paper	Prior genetic testing and expert feedback on manuscript		
Signature		Date	30/3/16

6.3 Manuscript

Self-reverting mutations partially correct the blood phenotype in a Diamond Blackfan Anemia patient

Parvathy Venugopal[1-3], Sarah Moore[1], David M Lawrence[3,4], Amee J George[5-7], Ross D Hannan[5-10], Sarah CE Bray[2,13], Luen Bik To[11,13], Richard J D'Andrea[2,11,12], Jinghua Feng[4,12] , Amanda Tirimacco[1], Alexandra L Yeoman[1], Chun Chun Young[1] , Miriam Fine[14], Andreas W Schreiber[3,4], Christopher N Hahn[1,2,13], Christopher Barnett[13,14], Ben Saxon[12, 15] and Hamish S Scott[1-3,12,13 #]

[1]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia; [2]Centre for Cancer Biology, SA Pathology, Adelaide, Australia; [3]School of Biological Sciences, University of Adelaide, SA 5005, Australia; [4]Australian Cancer Research Foundation Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, South Australia, Australia; [5]Oncogenic Signalling and Growth Control Program, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia; [6]John Curtin School of Medical Research, Australian National University, Acton, Australia; [7]School of Biomedical Sciences, University of Queensland, St. Lucia, Queensland, Australia; [8]Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Victoria, Australia; [9]Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria, Australia; [10]Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia; [11]Division of Haematology, SA Pathology, Adelaide, Australia; [12]School of Pharmacy and Medical Sciences, Division of Health Sciences, University of South Australia, Australia; [13]School of Medicine, University of Adelaide, Adelaide, South Australia, Australia; [14]South Australian Clinical Genetics Service, SA Pathology, Women's and Children's Hospital, North Adelaide, South Australia, Australia; [15]Department of Haematology, SA Pathology, Women's and Children's Hospital, North Adelaide, South Australia, Australia

Author for correspondence
Professor Hamish S. Scott
Department of Genetics and Molecular Pathology
Centre for Cancer Biology
SA Pathology
PO Box 14, Rundle Mall,
Adelaide, 5000, South Australia
Australia
Phone: +61 8 8222 3651
Fax: +61 8 8222 3146
Email: hamish.scott@health.sa.gov.au

Abstract

Diamond Blackfan Anemia (DBA) is an inherited bone marrow failure disorder associated with mutations/deletions in proteins involved in ribosome biogenesis. We have identified multiple underlying genetic defects in a complex case of DBA – a 184 kb deletion on chromosome 12 encompassing two ribosomal genes (*RPS26* and *RPL41*) and a heterozygous splice donor site mutation in *SBDS* which is implicated in Schwachman Diamond Syndrome (SDS), another inherited bone marrow failure disorder carrying significant clinical overlap with DBA. The DBA phenotype however, is most likely caused by haploinsufficiency of *RPS26*. Significantly, we also observed two independent copy neutral loss of heterozygosity events on chromosome 12, correcting the entire 184 kb deletion in a subset of blood cells, and propose this as a mechanism by which the patient's spontaneous remission occurred.

Diamond Blackfan Anemia (DBA) is a rare blood disorder characterized by red cell aplasia and is often accompanied by a variety of congenital abnormalities. It typically presents in infancy with macrocytic anemia and is associated with an increased risk of myelodysplastic syndrome (MDS), leukemia and solid tumors(1). The overall actuarial survival at 40 years is 75%. Current modes of therapy include corticosteroid treatment or red cell transfusions. Treatment-related complications like iron overload (treated by chelation therapy), infections and bone marrow (BM) transplant related complications are leading causes of death in most DBA patients rather than severe aplastic anemia and malignancy. Although the mechanisms remain unclear, approximately 20% of DBA patients enter spontaneous remission, wherein physiologically stable hemoglobin (Hb) levels are maintained in the absence of steroid therapy or transfusions(2).

Alongside other inherited marrow failure disorders with overlapping clinical features caused by ribosomal gene defects like Shwachman-Diamond syndrome (SDS) and dyskeratosis congenita (DC), DBA is classified as a ribosomopathy. Most cases of DBA are due to mutations in ribosomal proteins (RP) usually resulting in haploinsufficiency. Autosomal dominant inheritance of mutations in components of the 60S subunit (RPL5, RPL11, RPL15, RPL26, RPL35A) and 40S subunit (RPS7, RPS10, RPS17, RPS19, RPS24, RPS26) cause DBA while SDS follows autosomal recessive inheritance(3, 4). Interestingly, despite the ubiquitous requirement for ribosomes, germline mutations in RP cause tissue-specific defects, the mechanisms of which are unknown. We have used whole genome sequencing (WGS) and high density single nucleotide polymorphism (SNP) arrays to solve a complex case of DBA.

The proband ((II.3) Figure 1A), an 8-year-old boy (as of May 2016), is the third child of a Caucasian couple. He presented with anemia at 5 weeks (Hb 47g/L) which was treated with red cell transfusion. He required monthly transfusions for recurrent anemia and profound reticulocytopenia for the following year. BM biopsy was consistent with clinical diagnosis of DBA. He had mild bilateral ptosis and mild hypertelorism. No hepatosplenomegaly, renal abnormalities or bone abnormalities were seen on clinical examination. The siblings and parents had normal hematological findings. Clinical history is described in Figure 1B and hematological parameters at various time-points have been compiled in Supplementary Table 1. Along with anemia, he has intermittent neutropenia, monocytopenia and lymphopenia with normal lymphocyte subsets and immunoglobulin levels. He has significant growth delay, crossing height and weight centiles and now

tracking below the 1st centile. Endocrine and gastroenterological assessment is normal. He has some delays in neurocognitive development.

WGS analysis of the trio (proband, parents) was carried out to identify the causal mutation in the proband, an 8-year-old boy with DBA. Of 54 candidate variants, 25 were established as benign from the literature and 28 were variants of unknown significance. A maternally inherited splice donor site variant in the Shwachman-Bodian-Diamond-Syndrome (*SBDS*) gene was identified; c.258+1G>C (Supplementary Figure 1A, Supplementary Table 2). This mutation is predicted to cause a frameshift (p.C84Yfs*3)(5). The trio was also screened with the TruSight Cancer Panel and the maternal *SBDS* variant was confirmed in the proband. Sanger sequencing on samples from all family members confirmed the variant in the mother (I-2) and proband (II-3) as well as a sibling (II-1) who was born with a cleft palate (Supplementary Figure 1B). RT-PCR and western blot analysis revealed lower expression of *SBDS* in lymphoblastoid cell lines (LCL) established from peripheral blood (PB) in the family members with the variant compared to normal controls (Supplementary Figure 1C, 1D). However, we were unable to identify defects/imbalances pertaining to the paternal allele of *SBDS* using targeted next generation sequencing (NGS) on PB, BM and hair in the affected and hence continued to search for other causal mutated genes.

High density SNP microarrays identified a 184 kb deletion on chromosome 12q in the proband that was not present in either of the parents (Figure 2A). This was verified by a single read from WGS spanning the breakpoints of the 184 kb deletion. This deletion was also detected in the proband's hair sample, indicating a germline *de novo* mutation (Supplementary Figure 2B). The deleted region encompasses 11 genes (*PMEL, CDK2, RAB5B, SUOX, IKZF4, RPS26, ERBB3, PA2G4, RPL41, ZC3H10* and *ESYT1* - Supplementary Table 2). Heterozygous deletion or mutation of *RPS26*, a component of the small ribosomal subunit, has previously been described to cause DBA in an autosomal dominant manner(4). Intriguingly *RPL41*, which encodes a protein component of the large ribosomal subunit, also lies within the deleted region and has not been previously associated with DBA.

The 184 kb deletion breakpoints were further defined by analysis of read depths from WGS data (Supplementary Figure 3). PCR using primers flanking these coordinates on gDNA yielded a 3.7 kb product (Figure 2C) in II-3 and not in other family members. Sanger sequencing using nested primers revealed that the deletion was the result of a

homologous recombination event between Alu repeats (AluY: 56354900-56355009 and AluSc8: 56539460-56539754) (Supplementary Figure 4), which is not an uncommon mechanism contributing to germline disease(6).

Moreover, SNP microarray identified two additional regions of copy neutral loss of heterozygosity (cnLOH) on chromosome 12 in PB, LCL and BM, but not hair of the proband (Figure 2B; Supplementary Figure 2). The presence of multiple somatic clones and the tendency of DBA patients to have an increased risk of myeloid malignancy raised the possibility that the patient may be pre-leukemic. To investigate this, the blood sample was interrogated for recurrent mutations associated with leukemia by targeted NGS. When analysed in comparison with hair, it was found to be free of any predicted pathogenic somatic mutations down to a 5% variant allele frequency.

Allelic imbalance with an over-representation of the paternal allele was seen on chromosome 12 across SNPs where the affected child was heterozygous, in the targeted NGS data (Supplementary Table 3). This was consistent with a case of two independent revertant mutations wherein the 184 kb deletion defect in the maternal copy of chromosome 12q was being 'corrected' by replacement with the intact paternal copy. The patient's spontaneous reticulocyte recovery may be a consequence of the existence of two stably represented, non-identical clones exhibiting cnLOH spanning chromosome 12q. Absence of these cnLOH events in an earlier sample from the patient (prior to spontaneous recovery) lends further support to this hypothesis (Figure 2F). Also, the proportion of clones exhibiting cnLOH appears higher in LCL (Supplementary figure 2C) suggestive of a growth advantage. Erythroid colony assays from PB mononuclear cells revealed reduced colony forming efficiency (Supplementary Figure 5). Attempts to genotype informative SNPs were unsuccessful due to colonies being extremely small with possible contamination from surrounding cells.

Revertant mosaicism has been described in several genetic diseases including skin disorders, DC and Fanconi anemia(7-9), and spontaneous remission is not uncommon in DBA(2). Moreover, mosaicism has been reported in DBA patients who attained remission(10). However, in the event of remission, the risk of developing MDS or leukemia is likely to remain in proportion to the burden of any residual disease causing clone. The expansion of reversion events arising independently suggests a selective advantage over the defective cell population with likely contribution to correcting the phenotypic defect. We propose that revertant mosaicism is one of the mechanisms by which spontaneous remissions are attained in DBA. In such cases, where auto-correction

has occurred in the hematopoietic compartment, it may be necessary to utilize alternative sources of DNA such as hair or saliva, when testing to identify the causative genetic lesion.

The DBA phenotype in the proband can primarily be attributed to haploinsufficiency of *RPS26*, and the effect of defects in ribosomal genes *SBDS* and *RPL41* on disease phenotype and treatment outcome are currently unknown. We hypothesise that the patient's developmental delay is secondary to the *de novo* chromosome 12q deletion. Craniofacial structures are derived from neural crest cells and in animal models, mutations in RP can cause apoptosis of neural crest progenitors(11). Human ribosomopathies are often accompanied by craniofacial abnormalities(12). Though the proband has very mild craniofacial abnormalities, it is possible that the cleft palate in the sibling (II-2) is due to the heterozygous *SBDS* variant, especially considering the variable presentation of congenital abnormalities in individuals with biallelic mutations in *SBDS*(13). This complex molecular genetic diagnosis has been achieved by using multiple powerful technologies each with strengths and limitations. Together however, they have provided valuable insight into disease biology of not only DBA but likely other similar disorders.

Acknowledgements:

We wish to acknowledge the patient's family for their continuing support, cooperation and commitment to the cause of bone marrow failure research. We would like to thank Lynda Williams and Heather Thorne OAM from kConFab at Peter MacCallum Cancer Centre, Melbourne, Australia, for assistance with preparation of LCLs; and Amilia Wee for generation of initial reagents for this project. We would also like to acknowledge the Australian Diamond Blackfan Anaemia (ADBA) programme members, consisting of Sheren Al-Obaidi, Sarah CE Bray, Richard J D'Andrea, Jianmin Ding, Amee J George, Thomas J Gonda, S Peter Klinken, Piyush Madhamshettiwar, Lorena Núñez Villacis, Richard B Pearson, Ben Saxon, Hamish S Scott, Kaylene J Simpson, Adam Stephenson, Amilia Wee, Louise N Winteringham, Mei S Wong, and Ross D Hannan for their contribution to this work. This work was supported by an Australian Diamond Blackfan Anaemia (ADBA) programme grant from the Captain Courageous Foundation and by grant funding from the National Health and Medical Research Council, Australia (APP10024215, APP1023059) and Adelaide Scholarships International from the University of Adelaide. This work has been submitted in partial fulfilment of the requirement for a PhD degree at the University of Adelaide for PV.

Authorship

Contribution: PV, CNH, BS and HSS wrote the manuscript. CNH, RDH, LBT, RDA and HSS planned experiments. PV, AG, AT, AY, CCY and SCEB performed assays. PV, SM, DML, JF, AT, AWS, CNH carried out different aspects of data analysis. CB and BS contributed to clinical aspects of data analysis. All authors critically reviewed and approved the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Hamish S Scott, Department of Genetics and Molecular Pathology, Centre for Cancer Biology, SA Pathology, PO Box 14, Rundle Mall, Adelaide, 5000, South Australia, Australia; Email: hamish.scott@health.sa.gov.au

Supplementary information is available at Leukemia's website.

References

1. Vlachos A, Ball S, Dahl N, Alter BP, Sheth S, Ramenghi U, et al. Diagnosing and treating Diamond Blackfan anaemia: results of an international clinical consensus conference. British journal of haematology. 2008;142(6):859-76.
2. Vlachos A, Muir E. How I treat Diamond-Blackfan anemia. Blood. 2010;116(19):3715-23.
3. Gazda HT, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Schneider H, et al. Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. American journal of human genetics. 2008;83(6):769-80.
4. Doherty L, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Clinton C, et al. Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. American journal of human genetics. 2010;86(2):222-8.
5. Woloszynek JR, Rothbaum RJ, Rawls AS, Minx PJ, Wilson RK, Mason PJ, et al. Mutations of the SBDS gene are present in most patients with Shwachman-Diamond syndrome. Blood. 2004;104(12):3588-90.
6. Deininger PL, Batzer MA. Alu repeats and human disease. Molecular genetics and metabolism. 1999;67(3):183-93.
7. Jonkman MF, Pasmooij AM. Realm of revertant mosaicism expanding. The Journal of investigative dermatology. 2012;132(3 Pt 1):514-6.
8. Hamanoue S, Yagasaki H, Tsuruta T, Oda T, Yabe H, Yabe M, et al. Myeloid lineage-selective growth of revertant cells in Fanconi anaemia. British journal of haematology. 2006;132(5):630-5.
9. Jongmans MC, Verwiel ET, Heijdra Y, Vulliamy T, Kamping EJ, Hehir-Kwa JY, et al. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. American journal of human genetics. 2012;90(3):426-33.
10. Farrar JE, Vlachos A, Atsidaftos E, Carlson-Donohoe H, Markello TC, Arceci RJ, et al. Ribosomal protein gene deletions in Diamond-Blackfan anemia. Blood. 2011;118(26):6943-51.

11. Trainor PA, Merrill AE. Ribosome biogenesis in skeletal development and the pathogenesis of skeletal disorders. Biochimica et biophysica acta. 2014;1842(6):769-78.
12. Narla A, Ebert BL. Ribosomopathies: human disorders of ribosome dysfunction. Blood. 2010;115(16):3196-205.
13. Myers KC, Bolyard AA, Otto B, Wong TE, Jones AT, Harris RE, et al. Variable clinical presentation of Shwachman-Diamond syndrome: update from the North American Shwachman-Diamond Syndrome Registry. The Journal of pediatrics. 2014;164(4):866-70.

Figure Legends:

Figure 1. Pedigree and disease progression of the proband with DBA. (A) Pedigree describing genetic lesions in the proband with DBA and his parents and siblings who display normal hematological findings. (B) Timeline depicting progress of disease and treatment strategies for the proband. Karyotype was found to be normal when tested (corresponds to red arrows).

Figure 2. Multiple defects in chr 12q in the proband. (A) SNP microarray reveals 184 kb deletion encompassing 11 genes on chr 12q and (B) copy neutral loss of heterozygosity comprising two regions of chr 12q - denoted by green and red boxes. (C, D) PCR across 184 kb breakpoint in PB and LCL gives product of expected size in the proband which was then confirmed to be *de novo* germline in a hair sample (E). (F) SNP microarray reveals copy neutral loss of heterozygosity events were absent in the proband early in the disease prior to spontaneous recovery.

Figure 1.

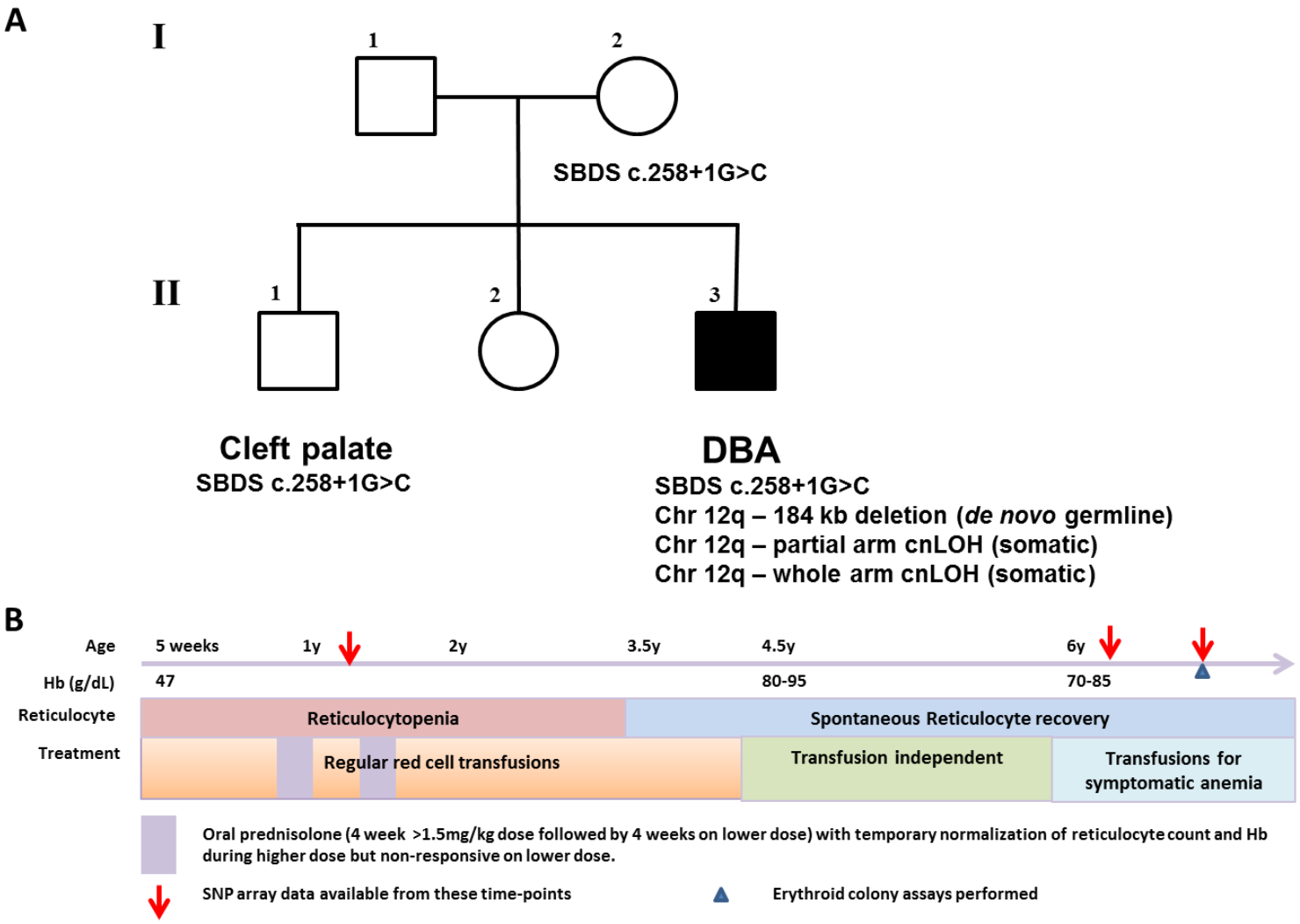
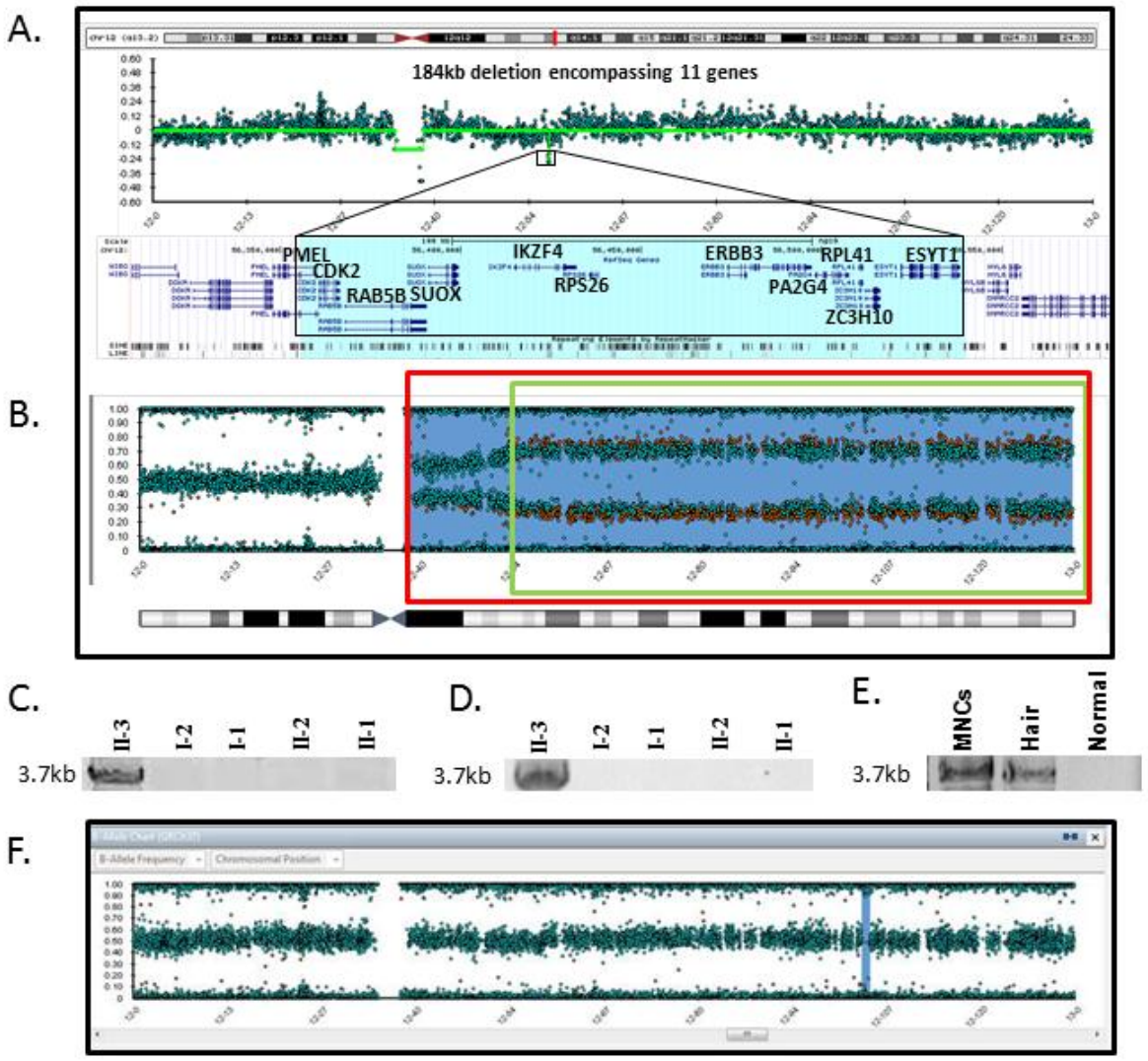

A
I
1
2
SBDS c.258+1G>C
II
1
2
3
Cleft palate
SBDS c.258+1G>C
DBA
SBDS c.258+1G>C
Chr 12q – 184 kb deletion (de novo germline)
Chr 12q – partial arm cnLOH (somatic)
Chr 12q – whole arm cnLOH (somatic)
B
Age
5 weeks
1y
2y
3.5y
4.5y
6y
Hb (g/dL)
47
80-95
70-85
Reticulocyte
Reticulocytopenia
Spontaneous Reticulocyte recovery
Treatment
Regular red cell transfusions
Transfusion independent
Transfusions for symptomatic anemia
Oral prednisolone (4 week >1.5mg/kg dose followed by 4 weeks on lower dose) with temporary normalization of reticulocyte count and Hb during higher dose but non-responsive on lower dose.
SNP array data available from these time-points
Erythroid colony assays performed

Figure 2.


A.
184kb deletion encompassing 11 genes
PMEL
CDK2
RAB5B
SUOX
IKZF4
RPS26
ERBB3
PA2G4
RPL41
ZC3H10
ESYT1
B.
C.
II-3
I-2
I-1
II-2
II-1
3.7kb
D.
II-3
I-2
I-1
II-2
II-1
3.7kb
E.
MNCs
Hair
Normal
3.7kb
F.

Chapter 7. Final Discussion

We have studied both inherited and acquired genetic lesions, and gene expression patterns that contribute to myeloid malignancies. The studies have given us valuable insights into specific aspects of malignant myeloid disease biology.

7.1 Functional relevance of GATA2 ZF2 mutations

We investigated mutations in ZF2 of GATA2 that were associated with specific clinical phenotypes. Contrary to the general assumption that all mutations in GATA2 lead to loss of normal function of the protein and therefore haploinsufficiency, we discovered that the mutations have unique effects on DNA binding, protein folding, interaction with binding partners and haematopoietic differentiation.

We found that the mutations which severely affected the structural integrity of the zinc finger and/or contact to DNA (gC373R, gR361L) also showed minimal binding to DNA and hence reduced transactivation. These mutations were found to be associated with Emberger syndrome (characterised by lymphoedema, immunodeficiency and predisposition to MDS/AML). This is consistent with earlier findings where haploinsufficiency of GATA2 has been to shown to result in Emberger syndrome. Mutants gT355del and gR396W showed a strong reduction in DNA binding affinity in EMSA as well as ITC. Based on this finding, we predict that these mutations also would predispose to Emberger syndrome.

The somatic mutation uniquely reported in CML blast crisis (sL359V) displayed slightly enhanced (or at least equal) DNA binding affinity and transactivation as WT GATA2. This may be reflective of the finding that the mutation does not significantly alter tertiary structure of ZF2. This mutant exhibited reduction in CFU-M colonies in the *in vitro* colony forming assays similar to the other mutants but was the only mutant that generated a significantly increased proportion of BFU-E colonies.

The remaining mutations (gT354M, sR362Q and gR398W) which were clinically associated with MDS/AML and/or immunodeficiency showed reduced DNA binding affinity and transactivation of GATA2-response elements. However, these mutants

displayed variability in their interaction with PU.1 and CEBPA (data not shown) which may contribute to the preferences in myeloid malignancy subtypes seen for some of the mutants. Another factor that may contribute to this variability is the socio-economic status of the patients and hence the spectrum of lifestyle related environmental stresses and infections to which they are exposed. Recurrent or persistent infections may lead to exhaustion of HSCs and onset of immunodeficiency-related symptoms prior to any myeloid malignancy.

7.2 *GATA2* haploinsufficiency does not affect *in vitro* mast cell differentiation

Previous studies have evaluated mast cell differentiation in the absence (*i.e.* complete knockout) of *GATA2* and shown that it is essential for the process. We investigated the effect of *GATA2* haploinsufficiency on *in vitro* mast cell differentiation and function. The cell recovery, immunophenotype, cytokine release and measure of degranulation was similar for WT and *GATA2* haploinsufficient bone marrow-derived mast cells suggesting that at least in these assays there is no difference.

7.3 Gene expression patterns used in conjunction with current risk stratification strategies can improve risk prediction at the time of AML diagnosis

We have used expression profiles of 85 genes in our cohort of AML patients to narrow down three genes that are highly predictive of outcome. When combined with the ELN risk stratification which is used by haematologists worldwide, it greatly improves risk prediction. This model can be used at diagnosis to aid treatment decisions where the adverse risk groups may be treated from the outset with more aggressive therapy like bone marrow transplantation with the aim of improving patient survival.

7.4 UPD for correction of a gene mutation is a mechanism for spontaneous remission in Diamond Blackfan Anaemia

In studying a DBA patient and trying to understand the underlying genetic cause of disease, we have uncovered a likely mechanism through which remission (seen in ~20% of patients) occurs in DBA patients. The patient had two underlying germline and/or *de novo* genetic lesions linked to his phenotype – a 180 kb deletion in chr12q

encompassing ribosomal genes *RPS26* and *RPL41*; and a heterozygous mutation in *SBDS* - compound heterozygous mutations of which cause a clinically similar ribosomopathy, Shwachman Diamond Syndrome (SDS). Interestingly, this patient also had two independent clones that exhibited UPD encompassing different lengths of chr12q where the 180 kb deletion was located. These post-zygotic events (confirmed by their absence in hair) resulted in two major clones where the lost ribosomal genes were restored by UPD of the paternal chr12q. We predicted that this would cause a correction of the DBA phenotype. Clinical notes, confirmed spontaneous reticulocyte recovery and a switch from a monthly transfusion program for recurrent anaemia to a transfusion-independent phase which was maintained for 18 months.

7.5 Conclusion

The findings presented in this thesis have addressed some unanswered biological questions pertaining to myelopoiesis and myeloid disease. We have a better understanding of how specific GATA2 mutations differ in the type of haematopoietic disease they generate. We have discovered links between gene mutations and expression, to prognosis in cancer with potential applications for determining treatment strategies for patients. We also have new insights into a possible mechanism by which haematopoietic cells 'repair' themselves to revert from a bone marrow failure phenotype to partial or complete recovery of haematopoiesis. Understanding the biology of disease undoubtedly brings us closer to effective therapeutic strategies.

Appendix I. Splice factor mutations and alternative splicing as drivers of haematopoietic malignancy.

Immunological Reviews

Christopher N. Hahn
Parvathy Venugopal
Hamish S. Scott
Devendra K. Hiwase

Splice factor mutations and alternative splicing as drivers of hematopoietic malignancy

Authors' addresses
Christopher N. Hahn[1,2,3,4], Parvathy Venugopal[1,2,4], Hamish S. Scott[1,2,4,5], Devendra K. Hiwase[1,3,6]
[1]Centre for Cancer Biology, SA Pathology, Adelaide, SA, Australia.
[2]Department of Molecular Pathology, SA Pathology, Adelaide, SA, Australia.
[3]School of Medicine, University of Adelaide, Adelaide, SA, Australia.
[4]Molecular and Biomedical Science, University of Adelaide, Adelaide, SA, Australia.
[5]University of South Australia, Adelaide, SA, Australia.
[6]Department of Haematology, SA Pathology, Adelaide, SA, Australia.

Correspondence to:
Christopher N. Hahn
Molecular Pathology Research Laboratory
SA Pathology
Frome Road
Adelaide, SA 5000, Australia
Tel.: +618 8222 3897
Fax: +618 8222 3146
e-mail: chris.hahn@health.sa.gov.au

Acknowledgements
Special thanks to Milena Babic for help with collation of references and our MDS mutation data. This work was supported by NHMRC project grant APP1024215 and Royal Adelaide Hospital Haematology private practice funding. The authors have no conflicts of interest to declare.

This article is part of a series of reviews covering Hematologic Malignancies appearing in Volume 263 of Immunological Reviews.

Immunological Reviews 2015
Vol. 263: 257–278

Immunological Reviews
0105-2896

Summary: Differential splicing contributes to the vast complexity of mRNA transcripts and protein isoforms that are necessary for cellular homeostasis and response to developmental cues and external signals. The hematopoietic system provides an exquisite example of this. Recently, discovery of mutations in components of the spliceosome in various hematopoietic malignancies (HMs) has led to an explosion in knowledge of the role of splicing and splice factors in HMs and other cancers. A better understanding of the mechanisms by which alternative splicing and aberrant splicing contributes to the leukemogenic process will enable more efficacious targeted approaches to tackle these often difficult to treat diseases. The clinical implications are only just starting to be realized with novel drug targets and therapeutic strategies open to exploitation for patient benefit.

Keywords: alternative splicing, splice factors, leukemogenesis, hematopoietic malignancy, hematopoiesis

Introduction

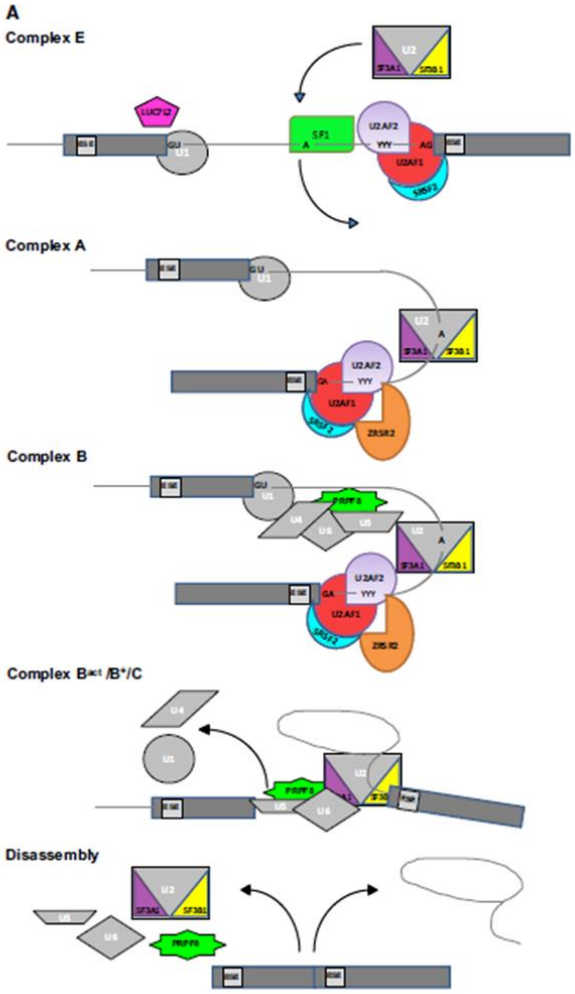
The splicing of nascent RNA transcripts into mRNA is a fundamental process in higher organisms that allows for the generation of multiple protein isoforms from the same gene as well as numerous non-coding transcripts. It has been estimated that alternative splicing (AS) from the approximately 20 000 protein coding genes can produce approximately 80 000 proteins (1) and probably many more (authors' unpublished data).

The modular structure of many proteins is in part generated by exon splicing. AS enables generation of a vast diversity of protein isoforms apart from the 'full length' isoform. These may be truncated, be missing one or more functional domains, or contain extra domains or new activities (neomorphic). AS variants may be (i) constitutively active, (ii) dominant negative, (iii) inactive, or (iv) able to interact with different partners/pathways. AS can also lead to changes in mRNA or protein stability, membrane-bound proteins becoming soluble, cell surface receptors changing their affinity for ligands or

co-receptors leading to aberrant signaling or migration, proteins being inappropriately sequestered in or trafficked to subcellular compartments, removal of autoregulatory domains, among many other possible changes (2–4). Splicing patterns are frequently cell-type specific, where inclusion or exclusion of particular exons are highly regulated. For some genes, this regulation acts as a rheostat where the ratio of alternative transcripts and protein isoforms reach particular thresholds that are crucial for cellular responses to stimuli or signals (5, 6).

Pre-mRNA splicing process

Splicing is a highly coordinated sequential process incorporating at least nine small nuclear ribonucleoproteins (snRNPs) and over 200 proteins that recognize 5′ splice sites (i.e. 5′SS or splice donor), 3′ splice sites (i.e. 3′SS or splice acceptor), or branch points, respond to exonic and intronic splice enhancers and silencers, and catalyze the splicing reaction (reviewed in 7–10) (*Fig.* 1). Positional, temporal, and magnitudinal information is encoded at and around exon-intron borders. Via protein:protein, protein:RNA, and RNA:RNA



B

Splice Factor	Function
SF3B1	- Core component of U2 snRNP -Targets U2snRNP to the branchpoint
SRSF2	- Exon recognition - Facilitates recruitment of both U2AF1 and SF3A2 at 3'SS
ZRSR2	- binds SR protein and U2AF2 - contacts 3'SS - essential for second step of splicing - capable of replacing U2AF complex in U12 dependent splicing
U2AF1	- binds to AG at 3'SS - recruits U2 snRNP - mediates interactions between U2AF2 and proteins bound to enhancers
PRPF8	- Binds U5 snRNP proteins - C-terminal is required for activation of Brr2 helicase unwinding of the U4/U6 duplex
LUC7L2	- recognition of non-consensus splice donor sites in association with the U1 snRNP
SF1	- Recognition of RNA branchpoint sequence (ACUNAC)
U2AF2	- Binds to polypyrimidine tract during early spliceosome assembly
SF3A1	- Necessary for activation of U2 snRNP in pre-mRNA splicing

Fig. 1. Spliceosome assembly and components mutated in hematologic malignancies. (A). Process of splicing. (i) U1 binds 5′SS; (ii) SF1 (branch point binding protein) interacts with the branch point sequence; (iii) U2AF complex recognizes the polypyrimidine tract and 3′SS via U2AF2 and U2AF1, respectively; (iv) U2 replaces SF1 at branch point site; (v) U4/U6 and U5 are recruited by U1 and U2; (vi) major rearrangements in RNA-RNA and RNA-protein interactions lead to removal of U1 and U4 from the complex; (vii) 5′SS and 3′SS are brought closer; (viii) two *trans*-esterification reactions cause intron loop/lariat formation and release of the lariat thereby joining the exons; (ix) components of the spliceosome disassemble for recycling. Colored elements represent components of the spliceosomal complex that are mutated in hematologic malignancies. (B). Splice factors with recurrent mutations in hematologic malignancies and their role in the splicing process.

interactions, the spliceosome precisely and reproducibly removes intronic RNA and generates mRNA. There are several excellent reviews detailing this process (7–10) (Fig. 1).

It is estimated that 92–94% of multi-exon genes in humans are alternatively spliced at least at some time during development or in some tissues (11, 12). The majority of introns are spliced by the major spliceosome known as the U2-dependent spliceosome which comprizes U1, U2, U4/U6, and U5 snRNPs. The U12-dependent spliceosome, which consists of U11, U12, U4atac and U6atac snRNPs, is involved in splicing of approximately 800 introns (U12 introns), which have a highly conserved 5′SS and branch point sequence and often lack the polypyrimidine tract (13). To date, U12 has not been associated with cancer.

The splicing process enhances the kinetics and efficiency of mRNA export from the nucleus and is linked to increased translational yield and genome stability. Nuclear export machinery proteins like UAP56 and NXF1 associate with the spliceosome and load onto mRNA when splicing occurs thereby contributing to improved efficiency of export. Intronless mRNAs require specific coding region sequences which confer stability and cytoplasmic localization. Increased translational yield is linked to deposition of exon junction complex (EJC) proteins on spliced mRNA (20–24 nt upstream of exon-exon junction). Some of these components remain bound until ribosomal passage during translation. Splicing is also essential for nonsense-mediated decay (NMD), an important RNA quality control mechanism. The EJC complex contains NMD factors (UPF proteins) that target transcripts with premature termination codons for degradation (14). Recently, it has been shown that mRNA splicing is regulated by BRCA1 in response to genomic instability, possibly to prevent transcription/translation of damaged genes and promote co-transcriptional splicing of a subset of genes for efficient DNA damage response (15). The activation of DNA damage response can affect AS by controlling expression, post-translational modification (through phosphorylation, acetylation, and ubiquitination) and intracellular localization of SF. In addition, AS can lead to different 5′ untranslated regions (5′UTR) or 3′UTR, impacting translation initiation/efficiency or miRNA-induced changes to mRNA stability/translatability, and ultimately protein abundance (16).

Alternative splicing or mis-splicing in cancer

The same mechanisms that enable normal cells to survive as they respond to a plethora of endogenous and exogenous signals and insults can be high-jacked in cancer, resulting in uncontrolled cell accumulation. It is widely accepted that changes in the transcriptome at both gene expression levels and AS play a crucial role in carcinogenesis in solid tumors (reviewed in 17–19). These changes may reflect the signature of a progenitor or stem cell of the particular cancer cell type (20) or may be due to aberrant splicing patterns set up by perturbations in spliceosomal components or genetic/epigenetic factors.

Global or more targeted changes in splicing impact on known cancer driver genes, and this is proposed to shift the balance of the cell to a more oncogenic state that reflects one or more of the 'hallmarks of cancer' (18, 20–22). There are many reports of global splice pattern changes in cancer. Some include colorectal, breast, and non-small cell lung cancer (20, 21, 23).

There are numerous examples of changes to single genes (e.g. expression levels, mutation, AS) that evoke an oncogenic-type response. Genes where changes in splicing have been shown to promote cell transformation include p53, CD44, FGFR2, and VEGFA. An AS variant of TP53 ('the guardian of the genome') (Δ133p53α), through interactions with p73, a cofactor that is required for TP53-dependent apoptotic cell death, is known to promote proliferation and tumorigenesis in gastric epithelial cells. Δ133p53α has the ability to regulate expression of canonical TP53(α) thereby impacting on its normal cellular function (24). CD44 is a cell surface protein involved in anchoring the epithelial cell to the extracellular matrix. Expression of a high molecular weight isoform of CD44 (CD44v6) has been associated with disease progression and poor prognosis in colorectal and ovarian carcinoma (25, 26). High CD44v6 is linked to increased mesenchymal markers, and hence may have a role in epithelial-mesenchymal transition (EMT), a process central to metastasis. In the progression of non-invasive to invasive breast cancer, a switch from FGFR2 IIIb to FGFR2 IIIc has been observed. This switch, like that for CD44, is associated with loss of epithelial markers and appearance of mesenchymal markers indicative of EMT (27). VEGFA exists in both pro-angiogenic and anti-angiogenic isoforms. In several solid tumors, the balance of expression is shifted to favor the pro-angiogenic isoform and this often correlates with tumor growth and poor prognosis (28).

One example of how AS can be regulated is the differential expression of the RNA binding protein Quaking (QKI). QKI is one of the most frequently downregulated SF in lung cancer tissue and is associated with poor prognosis (29). QKI has been to shown to specifically bind an ACUAAY motif which is very similar to the optimal binding sequence

for SF1 (ACUNAC; where underlined 'A' is the target of the lariat *trans*-esterification reaction), a core component of the spliceosome that binds to the intronic branch point (*Fig.* 1). The ACUAAY motif is enriched within 40 nt upstream of QKI repressed exons consistent with QKI competing with SF1 to regulate splicing. AS of NUMB [protein that promotes ubiquitination and degradation of NOTCH1 intracellular domain transcription factor (TF)] in cancer tissue can be correlated with QKI levels. One AS isoform of QKI (QKI-5) can compete with SF1 to inhibit inclusion of NUMB exon 12 leading to inhibition of NOTCH signaling and repression of downstream targets HEY2 and HES1 causing cell proliferation (29). In non-small cell lung cancer, AS was reported to increase inclusion of NUMB exon 9 leading to reduced NUMB protein levels, and derepression of NOTCH signaling to increase cell proliferation (21). High levels of NUMB exon 9 inclusion were also seen in colon and breast cancers (21). To add complexity, NOTCH1 acts as a tumor suppressor [e.g. neuroendocrine tumors, hepatocellular carcinoma, head and neck squamous cell carcinoma (30, 31)] or as an oncogene [e.g. T-cell acute lymphoblastic leukemia and chronic lymphocytic leukemia (CLL) (32, 33)] depending on cell type or cellular context. Remarkably, in the development of pancreatic ductal adenocarcinoma, NOTCH acts as a tumor suppressor early and as an oncogene in later stages (34).

These are only a few examples of ways in which AS can drive oncogenesis in solid tumors. We now focus on the role of splicing in normal hematopoiesis and HMs.

Alternative splicing or mis-splicing in HMs

AS is crucial for normal hematopoiesis (35). Key TF genes such as RUNX1, HOXA9, PU.1, and MLL (KMT2A) produce transcripts that are differentially expressed in various hematopoietic lineages (36–38). The presence, ratios, and stabilities of these AS transcripts and their subsequent protein isoforms are intricately controlled to achieve finely tuned homeostasis. Changes in AS transcript levels and hence the ratios of protein isoforms contributes to the leukemogenic phenotype.

RUNX1 is one such example. RUNX1 is a key hematopoietic TF that is often mutated in myeloid neoplasms where it most commonly acts as a true tumor suppressor gene. Loss-of-function mutation or deletion of one copy (haploinsufficiency) or complete ablation of its function promotes leukemogenesis. Furthermore, a minor splice isoform (1a) that lacks the transactivation domain and binds DNA with higher affinity than the full-length isoform (1b), acts in a dominant negative fashion, and is often seen at higher levels in acute myeloid leukemia (AML) (36).

HOXA9 is a homeobox protein that is highly expressed in hematopoietic stem cells (HSCs) and is downregulated during differentiation. It is often overexpressed in AML and ALL and is associated with poor prognosis (39). Recently, AS-derived isoforms of HOXA9 including natural isoforms lacking the DNA-binding domain (HOXA9T) transform myeloid cells to AML (40). Interestingly, HOXA9T that retains its ability to bind other HOX gene encoded proteins (MEIS1 and PBX1), does not act as a dominant negative but rather supports the leukemogenic potential of the full-length HOXA9 transcript. Hence, both full-length and a naturally occurring truncated isoform work in concert to drive leukemogenesis.

SPI1 (PU.1) is a key hematopoietic TF whose expression is intricately controlled in HSC and during differentiation of most blood cell lineages. SPI1 enforced expression causes transformation of proerythroblasts by inhibiting their differentiation while suppression of SPI1 is crucial for leukemic transformation of myeloid cells (41). SPI1 is a multi-faceted protein in that it interacts with numerous hematopoietic TF including GATA2 and GATA1, as well as spliceosomal proteins such as FUS (TLS) and NONO (42–44). Increased expression of SPI1 increases interaction with FUS, a SF that regulates AS, and blocks its splicing effects (45). More intriguingly, SPI1 impacts selection of specific exons, but only for genes, where it participates in the initiation of transcription. That is, in the process of SPI1 initiating transcription, it loads the RNA polymerase II complex with instructions that inform the transcribed pre-mRNA about choice of 5′SS, and hence the inclusion or exclusion of particular exons. The SPI1 DNA binding domain is required for both interference with FUS and promoter binding-dependent mechanisms of AS regulation (46). Hence, SPI1 affects AS via multiple strategies, and may contribute to leukemogenic processes.

The FLT3-STAT pathway controls many downstream target genes. One of these is the inhibitor of apoptosis gene BIRC5 (Survivin) (47), which regulates proliferation, cell division and apoptosis (48). Several novel BIRC5 splice forms have been described which associate with cytogenetics, leukocyte numbers and clinical responses in AML (49, 50). Furthermore, BIRC5 promotes abnormal proliferation in hematopoietic progenitor cells, and in the development of FLT3-mutated AML in a mouse model (51). In *de novo* pediatric AML patients, a high ratio of two minor splice variants (Survivin-2B:ΔEx2) correlates with poor outcomes

including overall survival (OS), event-free survival (EFS), and increased relapse risk (52).

WT1 (Wilms' tumor 1) is a TF affecting cell growth and differentiation. It is overexpressed in many leukemias and is mutated in some AML. The majority of mutations are heterozygous premature stops or missense both of which cluster in the C-terminus and are thought act in an oncogenic rather than tumor suppressor fashion. Over 30 different isoforms of the protein can be generated due to alternative transcription initiation, splicing, mRNA editing, and translation initiation (53). AS can lead to changes in DNA site recognition, DNA binding ability, and transactivation. Inclusion of WT1 exon 5 by AS is correlated with increased transactivation and progression from myelodysplastic syndrome (MDS) to AML (53).

Many of the BCL2 family of apoptosis genes generate AS variants. For some of these, changes in the ratio of pro- and anti-apoptotic isoforms determine cell survival (54). This has been well-documented for the $BCLX_L$ (anti-apoptotic) and $BCLX_S$ (pro-apoptotic) variants. Recently, AKT1, a major player in the phosphoinositide 3-kinase (PI3K)/AKT cell survival pathway, has been shown to be differentially post-translationally modified by SUMOylation, and that this modification increases the proportion of $BCLX_L$:$BCLX_S$ via AS, consistent with its pro-survival role. This function of AKT1 is independent of its well-studied phosphorylation by phosphoinositide-dependent protein kinase 1 (PDK1) (in response to PI3K activation) or mammalian target of rapamycin complex 2 (mTORC2) (55).

The BIM gene encodes up to 18 different AS variants and subsequent protein isoforms. An intronic deletion polymorphism common in the Asian population (minor allele frequency = 0.12) causes aberrant splicing of BIM to generate BIMγ that lacks a pro-apoptotic BH3 domain and inhibits apoptosis (56). This polymorphism confers resistance to tyrosine kinase inhibitors such as the imatinib in the treatment of chronic myeloid leukemia (CML). The SF SRSF1 contributes to splicing of BIM and another BCL2 family member, MCL1. Upregulation of SRSF1 in solid tumors promotes AS of transcripts from these two genes (57). Hence, changes to the ratio or abundance of variants from one or more apoptosis genes may impact on cell survival contributing to HM or success of treatment.

These are only a few examples of changes to natural splice variants or the introduction of aberrant splice variants that have a major impact on cellular processes associated with the 'hallmarks of cancer' (18, 58). We now focus on factors of the splicing process that when mutated play key roles in generating cellular, subcellular, or extracellular environments of leukemogenic potential.

Mutations in SF or spliceosome components contribute to HMs

In HMs and myeloproliferative neoplasms (MPNs), it has been known for some time that there is not only a shift in gene expression patterns but also a shift in splicing patterns. Examples include AML, MDS, and CML (53, 59–61). Only in the last 5 years have seminal discoveries demonstrated the significant and recurrent role of SF mutations in the pathogenesis of MDS, CLL, AML, and MDS/MPN. A select subset of SF is mutated, predominantly those associated with binding to the 3′SS or branch point. Presumably, this is because perturbation of 5′SS recognition is likely to cause intron retention and translation read through with severe consequences. Aberration of 3′SS recognition is more likely to cause exon skipping with less severe results. Recurrently mutated SF include SF3B1, U2AF1, SRSF2, and ZRSR2. Others that are rarely mutated include SF3A1, U2AF2, PRPF8, PRPF40B, SF1, SNRNP200, and LUC7L2 (each ≤1–2% of total SF mutations) (62–64). LUC7L2 is one of the few SF mutated that is associated with the 5′SS. In particular, it is proposed to be involved in recognition of non-consensus splice donor sites together with U1 snRNP (65). Notably, the majority of components of the splicing process have not been seen to be mutated, probably because mutations in these are either non-oncogenic or detrimental to the cell.

Why mutations in some components of the 3′SS and branch point recognition complexes occur much less frequently than others is not clear. As with mutations in any cancer-related gene, it is likely due to (i) a combination of types of mutational processes within the cell type, (ii) whether these mutations impact on amino acid residues that convey an appropriate 'oncogenic' change to the protein, (iii) level of penetrance of an oncogenic phenotype driven by a particular mutant gene, and (iv) whether gene mutations require preceding mutations or a permissive microenvironment to express their oncogenicity.

The frequency of the four most recurrent mutations varies depending on disease (Fig. 2). For instance, all four are seen in MDS, CMML, and AML, while only SF3B1 is common in CLL. For SF3B1, U2AF1, and SRSF2, nonsynonymous 'hotspot' mutations with an absence of nonsense and frameshift mutations are highly suggestive of gain-of-function mutations, possibly even neomorphic changes. ZRSR2 mutations are missense, frameshifts, and nonsense and

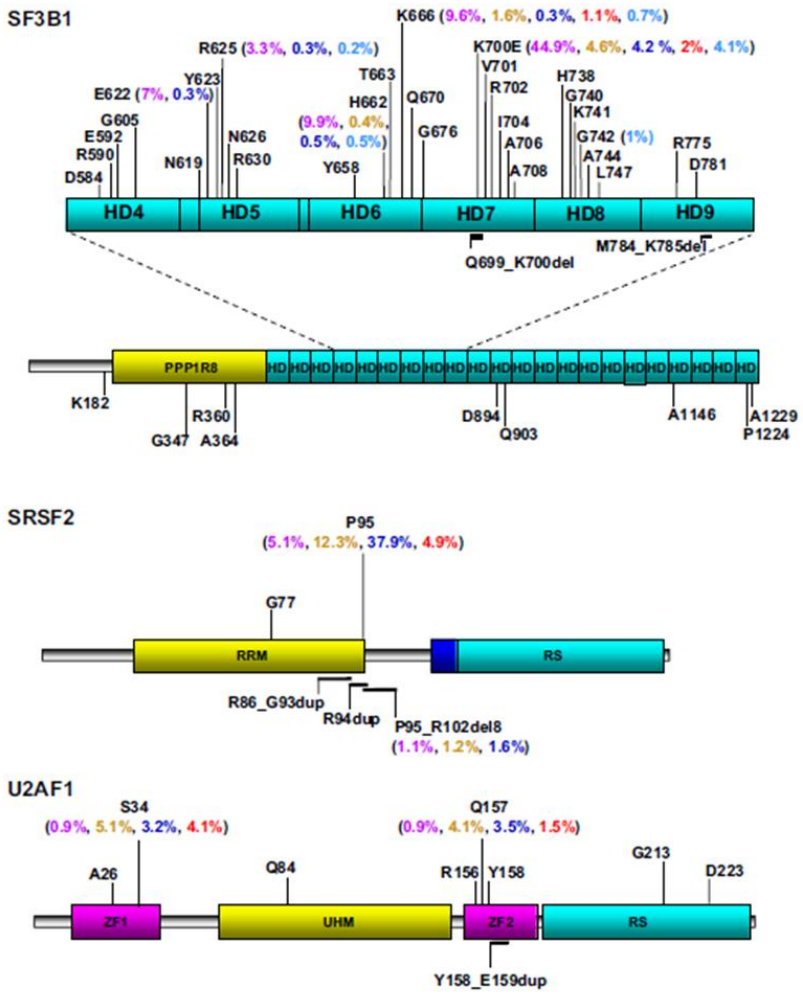


Fig. 2. Mutation of SF3B1, SRSF2, and U2AF1 in hematologic malignancies. Mutated residues and percentage of patients with mutations in that particular residue (brackets) are shown for RARS/RCMD-RS (purple), other MDS (yellow), CMML (blue), AML (red) and CLL (light blue). Binding site for protein phosphatase 1, regulatory subunit 8, PPP1R8; U2AF homology domain, UHM; arginine-serine rich domain, RS; RNA recognition motif, RRM (63, 68, 81, 88, 92, 97, 127–129, 132, 143, 157).

occur across the protein, indicative of a loss-of-function. Given that ZRSR2 is on the X-chromosome and that the vast majority of ZRSR2 mutations occur in males, this points to complete loss-of-function as the mode of its leukemogenic action.

It is becoming clear in HMs that mutations in epigenetic modifiers such as TET2, DNMT3A, and IDH1/2 promote self-renewal and block differentiation in the hematopoietic stem or progenitor cells giving them a growth, survival, and/or retention advantage (66, 67, authors' unpublished data). This leads to the bone marrow stem cell niche becoming clonally 'over-grown' with preleukemic stem cells while still allowing for the generation of a full repertoire of blood cells. A consequence of this is a heightened risk of developing malignancy and the subsequent persistence of these mutations into remission. Several studies looking at mutant allele frequencies, sub-clonal populations and progression samples have shown that SF mutations often occur early in tumor progression and are retained throughout disease evolution (62, 68–71). Therefore, like the epigenetic regulators, they too may be founder or initiator mutations or early events, and are required to maintain the oncogenicity or selectivity of the cancer cell (72). However, unlike the epigenetic modifiers, SF mutant preleukemic cells do not seem to persist with high allelic burden in remission (62, 68–71). While the mutations in these SF can impact disease

outcome, emerging oncogenic mutations may have a greater impact on patient outcomes (73).

One might expect that mutations in these general SF would cause extensive perturbations with devastating consequences. However, the splicing effects and disease outcomes are different for the 4 most commonly mutated SF. Below are some of the proposed mechanisms of action of these mutated SF.

SF3B1 is a critical component of both major (U2-dependent) and minor (U12-dependent) spliceosomes (74). It forms part of the U2 complex that displaces and replaces the branch point binding SF1 from the nascent transcript. The U2 complex is then involved in interactions with 5′SS and 3′SS components to facilitate intron lariat formation and excision (*Fig.* 1). Point mutations in the HEAT domain repeats (particularly repeats 5–7) or knockdown of protein levels cause different unique splicing aberrations, where mutation-induced AS changes are oncogenic while knockdown changes are not (75, 76). Mutations in SF3B1 are predicted to be 'mildly' damaging and do not fall on major conserved structural components of the HEAT repeats (64). Structural analysis of a U2 snRNP sub-complex containing SF3B1 revealed that the 22 tandem HEAT repeats wrap in an S-shape around the outer surface of the complex with HEAT domain 6 falling at the hinge of a shell-like structure (77). Hence, the mutations may impact on interactions of SF3B1 with other spliceosomal components.

SF3B1 mutations are clustered in five adjacent HEAT regions with K700E being most common and nonsynonymous mutations between E622-N626 and H662-Q666 constituting the majority of the remainder (78). While K700E is the most common SF3B1 hotspot mutation in MDS (58% of all SF3B1 mutations), CLL (50%) (63, 68, 69, 79–81) [15/20 (75%) SF3B1 positive MDS samples (authors' unpublished data)], and breast and pancreatic cancers (82, 83), mutations at R625 are most common in uveal melanomas of the eye (84, 85) and cutaneous melanomas (86). Because mutations at both of these sites are found in HMs and uveal melanomas, it is not clear whether this difference in predominant driver mutation is due to different mutational processes in the various cell types or due to subtly different selective forces that distinguish between the positioning of these mutations in the SF3B1 protein.

While mutations in several of the spliceosomal protein genes are seen in MDS, AML, and MDS/MPN (SF3B1, U2AF1, SRSF2, ZRSR2, SF3A1, U2AF2, PRPF8, PRPF40B, and SF1), SF3B1 seems to be uniquely mutated in CLL (79, 87). This observation suggests that disturbance of the splicing process by mutation in numerous SF is sufficient in myeloid malignancy, while specific requirement for mutated SF3B1 is found in CLL. In general, the range of SF3B1 mutations and their relative frequency of occurrence are similar for both malignancies. However, SF3B1 (G742D) has been recurrently seen in CLL (79, 81) but not found in approximately 1500 cases of MDS or myeloid neoplasms (63, 68, 73, 80, 88) (*Fig.* 2). Presumably there is a lymphoid specific factor or process affected by this change.

In MDS and MDS/MPN, SF3B1 mutations are highly linked to the formation of RS, abnormal red blood cell precursors characterized by aberrant accumulation of non-heme iron within mitochondria. SF3B1 mutation has a positive predictive value for RS of 97.7% (89) indicating a causal relationship. This is further supported by the strong correlation between mutant allele burden of SF3B1 mutations with the proportion RS (89). RNA-Seq analysis of SF3B1 mutant versus non-mutant refractory anemia with RS (RARS) patient samples identified AS of numerous genes including SLC25A37, an importer of iron into mitochondria (90, 91). AS of this gene resulted in two- to fourfold increased expression, providing a possible mechanism for the iron overload (91). Notably, SF3B1 mutations are common in CLL where RS do not occur. This is presumably because mitochondrial transporters, critical for the huge levels of heme required for myeloid (especially erythroid) development and hence RS phenotype, are not expressed in lymphoid lineages. In other mutant SF3B1 studies, AS was seen in genes critical for the pathology of MDS including ASXL1, RUNX1, EZH2, and CBL (92), all of which display oncogenic potential by gene disruption. Furthermore, significant changes were identified that impact on differential transcripts of QKI including QKI-5 isoform levels (see section on Alternative splicing or mis-splicing in cancer).

U2AF1 is a U2 auxiliary factor protein recognizing the yAG|r (i.e. T/CAG|G/A) intron|exon boundary at the 3′ SS (93). The 5′SS (GT dinucleotide) is recognized by the U1 RNA component of the U1 snRNP complex, while the 3′SS (AG dinucleotide) is recognized by the U2AF1 protein component of the U2 snRNP complex.

Virtually all oncogenic mutations are located at S34 and Q157 within the two highly conserved C_3H_1 zinc finger (ZF) domains of this RNA binding protein. The corresponding S34 amino acid in a highly related RNA binding protein ZFP36L2 directly hydrogen bonds with RNA while other aromatic amino acids of the protein stabilize binding to the pre-mRNA (94). The two recurrent substitutions, S34F and

S34Y, replace the smaller serine with larger aromatic residues, which is predicted to alter affinity for the RNA, possibly even increasing binding affinity. Some pre-mRNAs with strong polypyrimidine tracts are able to splice efficiently without U2AF1. Hence, intronic sequence determines the patterns of splicing of mutant U2AF1. One study showed mutant U2AF1 to promote enhanced splicing and exon skipping in reporter assays (95).

More recently, two groups studying AS patterns in U2AF1 wildtype versus S34F/Y tumor samples identified a select group of genes with a TAG 3′SS (underlined is the inerrant AG dinucleotide) where mutant U2AF1 aberrantly splices to a neighboring consensus or cryptic proximal or distal CAG or AAG 3′SS (96, 97). This results in inclusion or exclusion of entire or partial exons. Although widespread mis-splicing was not seen, the authors proposed that aberrant splicing may be more prevalent, but that NMD of such transcripts may 'hide' this phenomenon. Both groups identified several genes with known oncogenic activity in which AS correlated strongly with the presence of mutant U2AF1. These included CEP164, a centrosomal protein involved in G2/M checkpoint regulation and division of the nucleus, ATR, a protein in the DNA damage-signaling cascade that interacts with CEP164, and PTBP1, a factor that binds pyrimidine-rich RNA sequences and is known to influence AS possibly by competing with the binding of U2 snRNPs to nascent pre-mRNAs (97). AS of the RNA binding protein RBM10 was also seen in lung adenocarcinoma (96). Hence, for each of these genes and several other genes identified in cell cycle progression and RNA processing, their function can become oncogenic by somatic point mutation, deletion or mutant U2AF1-driven AS. These particular exon-skipping or mis-splicing events were unique to U2AF1 (S34F/Y) across 12 cancer types, and not seen in cancers harboring other SF mutations.

In contrast, Ilagan *et al.* (98) suggest that the effect of U2AF1 mutations are less selective and more widespread. They analyzed 169 U2AF1 mutant AML samples and the K562 cell line and, as above, confirmed that U2AF1 (S34F/Y) caused preferential exon skipping at TAG|r 3′SS and inclusion at CAG|r. Interestingly, the other recurrent Q157P/R mutations, unlike S34F/Y, equally skipped exons at yAG|G in favor of yAG|A. Knockdown of wildtype U2AF1 caused splice pattern changes, but no change in 3′SS selection. Therefore, mutations at S34 or Q157 or knockdown of protein levels cause different reproducible and unique splicing aberrations, where zinc finger mutation-induced AS changes are oncogenic while knockdown changes are not. Again, genes implicated in HM were affected including DNMT3B (DNA methylation) and ATR. Hence, U2AF1 mutations may contribute to leukemogenesis and indeed oncogenesis by generating widespread AS changes impacting on diverse molecular pathways.

SRSF2 is a member of the SR family of splicing regulators, so called because they contain a domain rich in serine (S) and arginine (R). SRSF2 missense mutations are highly recurrent at P95 or to a lesser extent at the surrounding amino acids (G93 to P107). Also small inframe indels have been reported, with P95_R102del8 being most recurrent in MDS and CMML (99) (*Table* 1). Structural modeling of the RNA recognition motif (RRM) of SRSF2 bound to an RNA 9-mer (AGCAGCGUA) revealed that Y92 contacts RNA. This amino acid resides close to the recurrently mutated P95 region suggesting that mutation alters interaction of SRSF2 with RNA (100).

ZRSR2 is structurally similar to U2AF1. It can participate in U2 and U12 spliceosomes and, during U12 splicing, it can replace the U2AF complex (*Fig.* 1). It may also stabilize the U2AF complex. Heterozygous mutations occur throughout the protein including frameshifts and nonsense mutations. Because ZRSR2 is on the X-chromosome and oncogenic mutations are almost exclusively seen in males, this is suggestive that complete loss-of-function plays an oncogenic role. Furthermore, since mutations are hemizygous, the allelic frequencies are often above 50% and can approach 100% in aggressive disease.

PRPF8 is the largest and most evolutionarily conserved protein in the spliceosome (101). It is a component of the catalytic core of the spliceosome and directly interacts with the 5′SS and 3′SS, the branch point, snRNAs, and the excised intron lariat (102, 103). It has been shown to function predominantly in the second step of splicing, acting as a scaffold and activating the BRR2 helicase to unwind the U4/U6 duplex for release of U4, an essential step for lariat formation (*Fig.* 1). PRPF8 missense, nonsense, and deletion mutations cause aberrant splicing of genes including some involved in hematopoiesis and iron metabolism in mitochondria (104). It is proposed that this may be due to defects in proof-reading leading to neomorphic spliceosomal activity possibly as a result of accumulation of non-functional spliceosomal complexes. The phenotype and splice patterns generated by mutations in PRPF8 are similar to but not identical to PRPF8 gene deletions (haploinsufficiency), implicating common critical aberrantly spliced driver genes in the generation of HMs. Reduced PRPF8 activity in the K562 erythroleukemia cell line and primary bone marrow

Table 1. Positional frequency of recurrent mutations of SF3B1, SRSF2, and U2AF1 in hematopoietic malignancies

Splicing Factor	Diagnosis (# patients)	% patients with SF mutation	% patients with Amino Acid mutated						
SF3B1			**K700**	**K666**	**H662**	**R625**	**E622**	**G742**	**Others**
	RARS/RCMD-RS (n = 272)	78.7	44.9	9.6	9.9	3.3	7.0	–	4.0
	Other MDS (n = 675)	7.6	4.6	1.6	0.4	–	–	–	0.9
	CMML (n = 384)	6	4.2	0.3	0.5	0.3	0.3	–	0.5
	AML (n = 452)	4	2.0	1.1	–	–	–	–	0.9
	CLL (n = 1749)	9.2	4.2	0.9	0.3	0.3	0.3	1.0	2.1
SRSF2			**P95**	**P95_R102del**	**Others**				
	RARS/RCMD-RS (n = 272)	6.3	5.1	1.1	–				
	Other MDS (n = 675)	13.8	12.3	1.2	0.3				
	CMML (n = 707)	40.2	37.9	2.0	0.3				
	AML (n = 452)	4.9	4.9		–				
U2AF1			**S34**	**Q157**	**Others**				
	RARS/RCMD-RS (n = 230)	1.7	0.9	0.9	–				
	Other MDS (n = 905)	9.3	5.1	4.1	0.1				
	CMML (n = 564)	7.3	3.2	3.5	0.5				
	AML (n = 784)	6.4	4.1	1.5	0.8				

Recurrent SF mutations in various hematopoietic malignancies were collated from various reports and their frequencies calculated (63, 68, 81, 88, 92, 97, 127–129, 132, 143, 157).

cells ($CD34^+$) led to increased proliferation and clonogenicity, hallmarks of many cancers. Like SF3B1 mutations, PRPF8 mutations are associated with accumulation of RS and explain some cases of wildtype SF3B1 MDS with RS.

Role of SF in non-splicing events

Although it is easy to focus on the roles of SF in splicing, they can be involved in other pathways. SF3B1 is implicated in numerous cancer-related pathways, including apoptosis, cell cycle control, and Hox gene regulation through both splicing and non-splicing mechanisms (105). $SF3B1^{(+/-)}$ mice displayed ectopic expression of Hox genes. These heterozygous mice remarkably had only 25% SF3B1 levels, but no reduction in splicing activity suggesting a non-splicing role for SF3B1. Repression of Hox genes by polycomb requires SF3B1 (87, 105). SRSF2 plays a role in genomic stability and senescence, seemingly independent of its splicing role (106–108). Depletion of SRSF2 has been reported to trigger overwhelming double-stand breaks. This is consistent with SRSF2 mutant MDS cases having significantly more mutations throughout the genome than is the case with other mutant spliceosomal genes (63) and may contribute to poorer patient outcomes (106, 109). SRSF2 also plays a role in transcriptional pausing of the RNA polymerase II complex thereby regulating rates of transcription (110). It activates transcription in a position-dependent manner by binding to promoter-associated small RNAs to mediate transcription pause release. Furthermore, several SR proteins enhance translation while some hnRNPs suppress mRNA levels in a position-independent fashion (110).

SF mutations and associated concurrent gene mutations

Mutations in SF such as SF3B1, U2AF1 and SRSF2 are typically early in the generation of MDS or AML (62, 68, 69, 95). They can occur before or after epigenetic mutations in genes such as DNMT3A, TET2, IDH1/2, and ASXL1 (based on clonal architecture) and may be the initial founder mutation. SF mutations tend to occur prior to mutations in TF and signaling molecules, possibly creating permissive environments. To date, germline mutations in SF have not been reported.

Unlike other HM driver genes such as RUNX1, TET2, and CEBPA, SF mutations are nearly always heterozygous, and biallelic mutations have not been reported. This implies either a dominant function of the mutant, an absolute requirement for some wildtype protein, or no additional selective advantage of a subsequent mutation. However, this does not preclude mutations in other SF contributing an additional selective advantage to a preleukemic or tumor cell. While several MDS and MDS/MPN studies have noted that mutations SF genes are largely mutually exclusive (i.e. providing an overlapping function or existing within the same pathway), most studies also report occasional (1–4%) concurrent mutations in different SF (62, 63, 68, 88, 97, 111, 112). This is consistent with the proposal by Mian *et al.* (73) that this is within a range of statistical expectation and that they are not mutually exclusive. Furthermore, Rozovski *et al.* (87) proposed that in most reports, the numbers of samples are not large enough to accurately determine whether infrequent mutations co-occur. On the other hand, they may be of significant size to demonstrate that SF mutations are not synergistically selected as are

rare biCEBPA and GATA2 mutations in AML (113). Indeed, different mutated SF show strong contrasting preference for co-occurring mutations, display 'cancer-specific' AS, and are found at different frequencies in different HM (68) (Fig. 2). This has led to a hypothesis where initiating mutations in different SF confer a 'genetic predestination' or permissivity that restricts the selective advantage of subsequent genetic changes to channel or confine the course(s) of clonal evolution (68). For instance, SF3B1 is causally linked to MDS with RS (>80%), while SRSF2 is the most frequently mutated spliceosome gene in CMML (68, 99). To date, there appears to be no evidence that subsequent mutations in different SF are not tolerated within the cell or are lethal. Hence, it is likely that mutations in different SF contribute non-overlapping advantages to the cancer cell. In CLL, concurrent SF mutations do not occur since only SF3B1 mutations have been found.

Some have assumed that the SF genes contribute to the same splicing pathway, but inferred that the downstream consequences to splicing must be different, and concluded that different combinations of comutated genes are instrumental in driving the phenotype of the disease (68). In MDS, concurrent SF mutations include SF3B1 with DNMT3A (111), SRSF2 with IDH1, IDH2, ASXL1 (112), RUNX1 (109), or STAG2 (112), U2AF1 with DNMT3A (88, 109), ASXL1 (88, 109), RUNX1 (112), BCOR (64, 92, 109) or chromosome 20 deletions (88), SRSF2 and ZRSR2 with TET2 mutation (88), and SF3B1 with deletion of chromosome 11q. Mutually exclusive mutated genes include SRSF2 with EZH2, and SF3B1 and SRSF2 with TP53 both at diagnosis and at the time of disease transformation (73).

Clinical outcomes of SF mutations

Spliceosome mutations in hematologic malignancies

Mutations involving multiple members of the mRNA splicing machinery including SF3B1, SRSF2, U2AF1 (U2AF35), ZRSR2, PRPF40B, U2AF2 (U2AF65), and SF1 are reported in patients with MDS, MPN, AML, and CLL.

Myelodysplastic syndromes

MDS are a heterogeneous group of clonal HSC malignancies which are characterized by ineffective hematopoiesis resulting in peripheral cytopenias, and typically a hypercellular bone marrow and an increased risk of progression to AML. Although the prevalence of MDS is not determined precisely, in the USA, >10 000 new cases of MDS are diagnosed annually (114). The World Health Organization (WHO) classification of tumors of hematopoietic and lymphoid tissues define the following categories: refractory cytopenia with unilineage dysplasia, refractory anemia with ring sideroblasts (RARS), refractory cytopenia with multilineage dysplasia (RCMD), refractory anemia with excess blasts (RAEB) type 1 (RAEB-1) and type 2 (RAEB-2), and MDS with isolated (del 5q) [MDS del(5q)] (115, 116).

Although cytopenias are a major clinical challenge in low-risk disease, transformation to AML is a major concern in high-risk MDS patients. The survival of MDS patients can range from 6 to 8 years in indolent conditions to <12 months in high risk groups (117, 118). There is increasing evidence indicating that chromosomal abnormalities play a major role in determining the heterogeneity of MDS (118, 119). However, conventional metaphase bone marrow cytogenetics is normal in almost 50% of MDS cases; hence until recently, the genetic aberrations that play an important role in the molecular pathogenesis of MDS was largely unknown. However, with the availability of next generation sequencing (NGS), genetic abnormalities have been reported in 80–90% of MDS cases (62, 120, 121). The most common mutations are detected in genes that are epigenetic modifiers (TET2, ASXL1, DNMT3A, EZH2, IDH1, IDH2), regulators of RNA splicing (SF3B1, SRSF2, U2AF1, ZRSR2), regulators of signal transduction (NRAS, KRAS, JAK2), and TFs (RUNX1, TP53) (63, 111, 120, 122). The majority of MDS patients harbor multiple mutations; 80%, 40%, and 10% MDS patients had ≥1, 2 to 3, and 4 to 8 mutations respectively (68). The mean number of mutations tends to be more in high risk subtypes (121). Except for SF3B1, DNMT3A, JAK2, and MPL mutations, the majority of common mutations occurred more frequently in high-risk WHO subtypes (RAEB-1/-2) than in RA/RARS subtypes (121).

Mutations of SF genes in MDS

Spliceosome mutations are the most common molecular abnormalities found in MDS, reported in 38–64% of all MDS cases (63, 64, 68, 73, 88, 92, 109, 121, 123), and are rare in (3.7%) non-MDS bone marrow failure syndrome cases (n = 107) (124). These mutations largely affect the core components of the initial steps, such as recognition of 3′SS of the pre-mRNA target intron (SRSF2 and U2AF1) or the recruitment of the U2 snRNP to the branch point proximal to the 3′SS that contains SF3B1. The surprisingly high frequency and specificity of mutations in this complex, together with their predominantly mutually exclusive occurrence unequivocally identified

compromised function of the E/A splicing complex are hallmarks of this unique category of myeloid neoplasms, playing a central role in the pathogenesis of myelodysplasia. The close relationship between the mutation types and unique disease subtypes also support their pivotal roles in MDS (88, 109). Of the four most commonly mutated spliceosome genes, mutations in SF3B1, SRSF2 and U2AF1 are heterozygous "hotspot" point mutations suggesting a gain-of-function while mutations in ZRSR2 likely lead to loss-of-function. Mutations in PRPF40B, SF1, SF3A1, and U2AF2 are rare in MDS; each occurs in approximately 1–2% of patients (63, 92) and are generally missense or nonsense mutations.

Mutations in SF3B1

SF3B1 is located at chromosomal band 2q33 and encodes a core component of the U2 small nuclear ribonucleoprotein complex (U2 snRNP). Mutations in SF3B1 were reported in 14–28% of MDS (64, 68, 73, 88, 89, 109, 125) and 19% of MDS/MPN cases (64, 88, 89, 109, 111). SF3B1 mutations were also reported in other hematologic malignancies, including primary myelofibrosis (PMF) (4%) (126), therapy-related AML, *de novo* AML (5%), essential thrombocythemia (ET) (3%), CLL (5%), multiple myeloma (3%), and chronic myelomonocytic leukemia (CMML) (5%). SF3B1 mutations were also found in breast cancer (1%), renal cancer (3%), and adenoid cystic carcinoma (4%) (64).

SF3B1 mutated patients present with a distinct clinical and morphological phenotype associated with the presence of RS and more pronounced defects in the erythroid lineage. RS are erythroblasts characterized by accumulation of aberrant mitochondrial ferritin, visualized by Prussian blue staining as a perinuclear ring of blue granules. The presence of ≥15% RS constitutes the diagnosis of myelodysplastic syndromes with RS (MDS-RS), which includes RARS and RCMD-RS and RARS with marked thrombocytosis (RARS-T). Note that RCMD-RS was recognized as a separate entity in WHO-2001 classification system, but was grouped with RCMD in the revised classification (2008). The frequency of SF3B1 mutation is significantly higher in patients with RARS (64–91%), RCMD-RS (37–76%) (63, 64, 80, 99, 123) and RARS-T patients (66%) compared to other WHO subtypes of MDS (3–13%) (63, 64, 89, 99, 111, 123). In an analysis of the bone marrow of non-RS MDS cases (n = 222), re-examination of bone marrow iron stain detected variable numbers of RS (1–74%) in 35% (74/222) cases. The RS were <15% (below the threshold of definition of RS) in 60% (44/74) cases while in 40% (30/74) cases RS were >15%. In this study, 91% of SF3B1 mutant cases had >15% RS, while 7% had 1–14% RS and only 3% patients did not have RS in their bone marrow. This suggests that the positive predictive value of SF3B1 mutation for disease phenotype with RS is 97% and the absence of RS had negative predictive value of 98%. This study also reported a significant relationship between SF3B1 mutant ($SF3B1^{Mut}$) allele burden and percent RS and bone marrow erythroblasts (89). The proportion of patients with RARS, RARS-T, and RCMD-RS were higher when SF3B1 mutant allele burden was ≥25% than those with mutant allele burden <25%. SRSF2 and ZRSR2 mutation has been described in approximately 7% of MDS-RS with wildtype SF3B1 ($SF3B1^{WT}$) (63).

Clinical correlation with SF3B1 mutation showed that patients with SF3B1 mutations have significantly lower hemoglobin, higher RBC-transfusion dependency, higher white blood cell and platelet counts, marked erythroid hyperplasia, and lower bone marrow blasts compared to $SF3B1^{WT}$ patients (64, 88, 89, 125). The majority of $SF3B1^{Mut}$ patients have normal cytogenetics and low/int-1 IPSS risk groups. However, some studies did not report such correlations (99, 125), which may be due to a variability in patient population, time point in disease course, previous treatment and/or sensitivity of mutation assays. The impact of SF3B1 mutation on disease outcome is also debatable. Some studies reported significantly higher leukemia free survival (LFS), EFS, and OS in patients harboring SF3B1 mutation even after controlling for other established risk factors such as age, gender, hemoglobin levels, platelet counts, cytogenetic risk, bone marrow blast percentage, IPSS, WPSS, and RS (64, 68, 73, 89, 92, 123). However, other studies did not find a survival advantage or time to leukemic progression in mutant SF3B1 cases compared to $SF3B1^{WT}$ (88, 99, 109, 125). The good prognosis of SF3B1 mutations could be due to its significant association with the good prognosis subgroup of MDS (i.e. RARS) rather than the mutation itself. In support of the latter, in a subgroup analysis of RARS and RCMD-RS cases, SF3B1 mutation status was not associated with superior OS or LFS (89). SF3B1 mutations displayed no impact on outcomes in PMF (126). Conversely, SF3B1 mutation correlated with poor OS in CLL patients (127, 128).

SF3B1 mutations are heterozygous missense mutations with an average mutant allele burden of 41% (5–70%) with very few cases of low mutant allelic burden. The median burden was similar in purified $CD34^{+}$ cell, granulocyte and bone marrow DNA (40%, 40%, 35%) (89). This suggests

that these mutations are acquired in hematopoietic stem or progenitor cells and are transmitted to their myeloid progenitors (89, 129). Hence, hemopoiesis is sustained by the dominant mutant clone and these somatic mutations are an early pathogenetic event. The mutations are frequently reported in HEAT domains 5–7, which is implicated in snRNP stabilization within the U2 snRNP complex of the major spliceosome (130). The most common recurrent SF3B1 mutation affects amino acid K700 (45–68% of reported SF3B1 mutated cases) (64, 73, 88, 89, 109, 125) followed by H662 (10%), K666 (10%), E622 (7%), D781 and R625 (6%) (64, 73, 89, 125). There was no significant difference among different mutations in regards to WHO subtype, clinical features and other hematologic parameters (89).

Mutations in SRSF2

SRSF2 mutation is more frequently seen in the MDS/MPN group (24%), particularly in patients with CMML (28–50%) (92) than in MDS (10–15%) (68, 73, 109, 111, 131). The majority of SRSF2 mutations are heterozygous, missense mutations with an average allele burden of 37%. SRSF2, codon P95 was most frequently mutated [80–96% of SRSF2 mutated cases (73, 109, 125)] leading to P95H/L/R (88, 109). A small number of patients have point mutations causing P96L or frameshift mutations in codon P95 (73, 88) and Y93fs*121 (73); whether these are pathogenic or passengers is unclear.

Unlike SF3B1, there was no association between SRSF2 mutation status and RS. Correlation of SRSF2 mutations with hematologic parameters is an issue of ongoing debate; some studies reported higher neutrophil counts (73), similar platelet counts (73, 109), higher hemoglobin levels (73), and lower transfusion dependency in $SRSF2^{Mut}$ compared with $SRSF2^{WT}$ cases. By contrast, other studies reported significantly lower neutrophils (88), higher frequency of thrombocytopenia (88), similar hemoglobin levels (88, 109), and similar RBC transfusion dependency (73, 109). Similarly, the impact of SRSF2 mutation on survival outcomes is controversial. SRSF2 mutation has been reported as an independent unfavorable prognostic factor for OS (68, 92, 109) and AML transformation (73, 109, 132) even in IPSS-Low and Int-1 risk group (92). Importantly, these driver mutations displayed equivalent prognostic significance, whether clonal or subclonal (68), probably since a subclonal population can become clonal over time. In other studies, however, OS and AML progression rate were not different between $SRSF2^{Mut}$ and $SRSF2^{WT}$ patients (88, 92).

Mutations in U2AF1

U2AF1 mutations are detected in 5–16% of MDS cases (63, 88, 109, 111, 129). Mutations in U2AF1 exclusively involve two highly conserved amino acid positions (S34F and Q157P/R/H) within the amino-and carboxyl-terminal zinc finger motifs flanking the U2AF homology motif (UHM) domain (63, 88, 109) (*Fig.* 2).

The majority of patients (80–90%) with U2AF1 mutations were males (73, 88). There was no significant difference in age (109), platelet (73), neutrophil counts (73), WHO classification (73, 109), IPSS-based karyotype (109), BM blasts (109), hemoglobin (109), IPSS (73, 109), or transfusion dependency (73, 109) between $U2AF1^{Mut}$ and $U2AF1^{WT}$ (109). In contrast with SF3B1 mutation, there was no association between RS and U2AF1 mutations. U2AF1 mutations were frequent in high-risk MDS/AML cohort (11%). The prognostic impact of U2AF1 is also debatable. Some studies reported mutations in U2AF1 had no impact on OS (73, 88, 109) and progression risk to AML while others reported an inferior OS (92) and higher progression rates to AML (129).

Mutations in ZRSR2

The frequency of ZRSF2 mutations varied between 3% and 11% of MDS patients (63, 88, 109, 111) and are more frequent in advanced cases. In contrast with SF3B1, SRSF2, and U2AF1, mutations in ZRSR2 are spread over the entire coding regions and the majority of mutations are nonsense, frameshift or involve 5′SS/3′SS that causes either a premature truncation or large structural change of the protein, leading to loss-of-function (63). ZRSR2 mutations have a higher mutant allelic burden of 60%. In some studies, patients harboring ZRSR2 mutations were almost exclusively males (24/25; 96%) (63, 88), while other studies did not report a similar observation. Both high allelic burden and male predominance are indicative of ZRSR2 being on the X-chromosome. Patients with ZRSR2 mutations often presented with isolated neutropenia, tended to have advanced disease with a majority being RAEB-1 and RAEB-2, and IPSS-Int-1 and Int-2 risk categories (88). The OS was not significantly different between $ZRSR2^{Mut}$ and $ZRSR2^{WT}$ cases (88, 109); however, AML progression rates were higher in $ZRSR2^{Mut}$ IPSS-Low and Int-1 risk group patients (88).

Association of spliceosome mutations with other mutations

Mutations in genes involved in the RNA splicing machinery are mostly mutually exclusive (63, 68, 109). However, 75% patients with spliceosome mutations have mutations in epigenetic modifiers and/or cell signaling/TF while only 25% of patients have isolated SF mutations (73).

Mutations in genes involved in RNA splicing and DNA methylation occur early, whereas driver mutations in genes involved in chromatin modification and signaling often occur later (68). Mutations in each SF tend to co-occur with particular genes involved in epigenetic regulation. Patients with SF3B1 are more likely to have DNMT3A (88, 121) or CEBPA (73) mutations, while SRSF2 mutations are associated with mutations in TET2 (73, 88, 121), RUNX1 (88, 109), IDH1 (109), IDH2 (121), NRAS (73), and FLT3 (73). U2AF1 mutations are associated with ASXL1 (88, 109), RUNX1 (62), and DNMT3A (88, 109) mutations. MDS patients with SRSF2, U2AF1, and ZRSR2 tend to have significantly more mutations in other genes than patients with SF3B1 mutations. Conversely, EZH2 and SRSF2, and IDH2 and SF3B1 have not been seen to co-occur (68).

The average mutant allele burden for SF3B1 and SRSF2 mutation is 41% and 38% respectively (73). The average mutant allele burden of co-existing point mutations presenting alongside these SF mutations is 35–37%, with mutant allele burden of epigenetic modifiers being higher (38–39%) than cell signaling/transcription regulator mutations (27–28%). In contrast, U2AF1 mutations had a lower average allele burden (26%) with the mutation burden of other co-existing mutations averaging approximately 39% comprising cell signaling/transcription regulator mutations (45%) and epigenetic modifiers mutations (33%) (73).

The patients harboring mutations in SF and genes in cell signaling/transcription regulation or TP53 had an extremely poor OS and progression free survival (PFS) compared to patients with spliceosome mutations without mutations in cell signaling/transcription regulation genes (73). In $SF3B1^{Mut}$ patients, RUNX1 mutation was associated with poor prognosis (123) while in $SF3B1^{WT}$ patients, the presence of TP53, DNMT3A or ASXL1 mutation was associated with poor OS and higher rate of AML transformation (88, 111). Similarly, the prognostic impact of SRSF2 and ZRSR2 mutations depends on the mutation status of TET2. $ZRSR2^{Mut}/TET2^{WT}$ is an independent unfavorable prognostic factor for OS and higher risk of AML transformation (88).

These data suggest that RNA splicing is one of the most commonly mutated pathway in MDS, and mutations in SF most probably occur early in disease initiation or evolution. The exact mutations play a major role in determining the clinical features of the disease and may also influence the subsequent genomic evolution of the disease, as patterns of co-operating mutations are significantly different between SF3B1 and SRSF2.

Functional impact of SF mutation

Despite a strong correlation between spliceosomal mutations and MDS, MPN, and CLL, the mechanisms by which these mutations perturb the process of RNA splicing and subsequently lead to disease are still unknown. Mutant U2AF1 promotes enhanced splicing and exon skipping in reporter assays *in vitro* (129). This novel, recurrent mutation in U2AF1 implicates altered pre-mRNA splicing as a potential mechanism for MDS pathogenesis. Analysis of transfected HeLa and TF1 cells with $U2AF1^{WT}$ or $U2AF1^{Mut}$ showed a significant enrichment of genes in the nonsense-mediated mRNA decay (NMD) pathway (63). NMD activation by $U2AF1^{Mut}$ was suppressed significantly by co-expression of $U2AF1^{WT}$, suggesting that the mutant protein likely inhibits function of the wildtype protein. Hence, mutant U2AF1 induces abnormal RNA splicing leading to the generation of aberrantly spliced RNA species with premature stop codons that induce NMD activity (63). Furthermore, $U2AF1^{Mut}$ significantly reduced cell proliferation, markedly increased the G2/M fraction and enhanced apoptosis (63), while ZRSR2 knockdown reduced cell viability. When murine HSC were retrovirally transduced with $U2AF1^{WT}$ or $U2AF1^{Mut}$ and transplanted into mice using a competitive reconstitution assay, engraftment capacity of hematopoietic stem/progenitor cells expressing $U2AF1^{Mut}$ was significantly reduced. This suggests that these mutants lead to loss-of-function of U2AF1 (63) most probably by acting in a dominant-negative fashion to the wildtype protein. SRSF2 is involved in regulation of DNA stability and depletion of SRSF2 can lead to genomic instability (106). Consistent with this, irrespective of the disease subtype, samples with SRSF2 mutations (not other SF mutations) had more mutations in other genes (63).

SF3B1 mutations could influence either splicing itself or interactions with the transcriptional complex. Gene expression profiling of $CD34^{+}$ cells from patients with SF3B1 mutations showed downregulation of genes involved in the mitochondrial ribosome and the electron transport chain (64). SF3B1 mutations lead to iron accumulation by affect-

ing key iron trafficking pathways. The mRNA level of SLC25A37 (mitoferrin-1), a protein involved in the mitochondrial iron delivery in erythroid cells was significantly overexpressed in SF3B1 mutants compared with both $SF3B1^{WT}$ RARS/RARS-T and healthy donors (91). $CD34^{+}$ cells from RARS patients also overexpressed other mitochondria-related genes particularly those involved in heme synthesis (e.g. ALAS2) (133) and reduced expression of ABCB7, a gene encoding a protein involved in iron transport from mitochondria to cytoplasm (134, 135). The expression level of ABCB7 correlated with percentage of marrow RS (134, 135) and forced expression restored erythroid growth and survival of RARS progenitors, while decreasing aberrant mitochondrial ferritin. ABCB7 is neither mutated nor methylated in MDS-RARS. However, in RARS patients harboring SF3B1 mutation, ABCB7 exon usage differed to normal bone marrow (136). A transient downregulation of SF3B1 during erythroid differentiation results in reduced expression of ABCB7 suggesting a link between SF3B1 and ABCB7 (136). Patients carrying SF3B1 somatic mutations have inappropriately low levels of hepcidin which may cause excessive reticuloenothelial iron release and parenchymal iron overload. Hence, multiple mechanisms may contribute to abnormal iron accumulation. This has implications in terms of management of iron overload, a debatable area in MDS.

Myelodysplastic/myeloproliferative neoplasm (MDS/MPN overlap)

WHO classification of tumors of hematopoietic and lymphoid tissue (2008) grouped chronic myelomonocytic leukemia (CMML), atypical CML (BCR-ABL1 negative), juvenile myelomonocytic leukemia (JMML) and MDS/MPN-unclassifiable under the heading of MDS/MPN, as they share features of both MDS and MPN.

Chronic myelomonocytic leukemia

CMML is a clonal HSC disorder characterized by features overlapping between MDS and MPN. The classification of CMML has changed multiple times, in the French-American-British Group (FAB) classification CMML was considered as a subtype of MDS due to dysplastic features, cytopenia and increased risk of AML progression. However, it became clearer that CMML is clinically distinct from MDS in 2001 when CMML was classified as a provisional category of MDS/MPN-overlap syndrome. In 2008, the WHO formalized the definition of CMML. CMML is a heterogeneous disease with poor long-term survival (137, 138). In approximately 15–30% of patients with CMML, the disease evolves into AML (137, 138).

Multiple prognostic scoring systems have been developed to predict CMML prognosis, but most are based on blood counts, bone marrow blast, cytogenetics, and age (117, 118, 138–141). With the introduction of NGS technology, mutations can be detected in >90% of CMML patients involving epigenetic regulator genes (EZH2, ASXL1, TET2, DNMT3A, IDH1, IDH2), spliceosome complex genes (SF3B1, SRSF2, U2AF1, ZRSR2, SF3A1, PRPF40B, U2AF2, SF1), genes regulating cellular/receptor tyrosine kinases and TF (JAK2, KRAS, NRAS, RUNX1) and DNA damage response genes such as TP53 (131). TET2, SRSF2, and ASXL1 mutations occur at particularly higher frequencies in CMML compared with other myeloid malignancies, but none of these alterations is unique to CMML.

Spliceosome mutations mainly in SF3B1, SRSF2, and U2AF1 are detected in 50–58% of CMML patients (63, 142–146) and are mutually exclusive (123). In contrast with phenotypic convergence, the mutational spectrum of JMML and CMML is divergent with mutations in spliceosome complex genes being very rare in children with MDS and JMML (142, 147), suggesting that these are quite different diseases.

In CMML, SRSF2 is the most commonly mutated spliceosome gene (28–47%) (63, 123, 131, 142–146, 148). The majority of SRSF2 mutations are heterozygous, missense mutations most frequently affecting codon P95 (93% of SRSF2 mutations) leading to P95H/L/R (144, 145). A small number of patients have inframe indels around P95 (Fig. 2). Patients with SRSF2 mutation were older and had higher hemoglobin levels and diploid karyotype compared to $SRSF2^{WT}$ cases. The white blood cell and platelet counts were not different between the two groups. SF3B1 mutations were detected in 4.5–6.5% of CMML (63, 89, 142, 143, 145, 146) with K700E, K666Q/N/R and E662D being predominant (142, 143, 145). U2AF1 mutations were detected in 7–13% of patients (142, 143, 145, 146). Most of the mutations were heterozygous, missense mutations Q157P and S34F (143, 145) (Fig. 2). Some cases (5/11) had concomitant LOH of 7q (142). Serial studies showed that U2AF1 and SF3B1 mutations were seen at the earliest presentation of the disease and their frequency did not increase significantly either serially or cross-sectionally when different disease stages were compared (142–144). More advanced stages of disease (CMML1 versus CMML2 versus sAML) were not associated with a higher mutational burden (142) suggesting the mutation is required early in disease development.

The presence of SRSF2, U2AF1 (or DNMT3A) mutations conferred an inferior OS (131, 142), while other studies reported similar OS and LFS in cases with SRSF2, SF3B1, and U2AF1 mutations and patients without spliceosomal mutations (142, 143, 145). Patients with both RUNX1 and SRSF2 (P95H) mutation had better survival compared to patients with $RUNX1^{Mut}/SRSF2^{WT}$ (144). The mutational rates of SRSF2, SF3B1, U2AF1, RUNX1, and CBL were similar in responders and non-responders to hypomethylating therapy (142). Spliceosomal gene mutations were mutually exclusive in most of the cases but were frequently associated with other non-spliceosomal gene mutations. SRSF2 mutation was commonly co-mutated with TET2 (131, 142, 144), ASXL1, RUNX1, and CBL (142), while mutations in SRSF2 and EZH2 genes were mutually exclusive (144). A co-occurrence of mutations in SF3B1 and in JAK2, MPL, or CALR was strongly associated with a MDS/MPN phenotype characterized by thrombocytosis (123). Patients with concurrent mutations of TET2 and SRSF2 or ZRSR2, showed significantly higher hemoglobin levels and higher monocyte counts compared to those without TET2 co-mutation.

Refractory anemia with ring sideroblasts and thrombocytosis

Refractory anemia with ring sideroblasts and thrombocytosis (RARS-T) was proposed in the WHO 2001 classification of tumors of hematopoietic and lymphoid tissues and retained as a provisional entity, under MDS/MPN overlap-unclassifiable group, in the revised WHO classification in 2008 (115). RARS-T has dysplastic features of RARS and the myeloproliferative features of ET. JAK2 (V617F) and/or MPL (W515K/L) mutations were detected in 40–77% cases of RARS-T (149–151), an incidence similar to that found in ET and PMF. The classification of RARS-T remains an active area of debate, where some specialists believe that RARS-T is a form of ET with >15% of RS while others believe that it develops from RARS with secondary thrombocytosis due to acquisition of JAK2 (V617F) mutation.

SF3B1 mutation is reported in 67–90% cases of RARS-T (80, 89, 151) with a median mutant allele burden of 41% (15–70%) (89, 151). The most frequent mutations reported were K700E (51%), K666R/N/T (16%), H662Q (9%), R625K/C (7%) (*Fig.* 2). In RARS-T, SF3B1 mutations were more frequent in females (95%) than in males (77%) and the mean RS counts were higher in $SF3B1^{Mut}$ than in $SF3B1^{WT}$. Apart from this, there was no difference in clinical and hematologic variables between $SF3B1^{Mut}$ and $SF3B1^{WT}$ RARS-T patients. The median survival of RARS-T patients with $SF3B1^{Mut}$ was significantly higher than patients without $SF3B1^{Mut}$ (6.9 versus 3.3 years, $P = 0.003$) (151). JAK2 (V617F) and SF3B1 mutations were detected in 43–64% of RARS-T cases (80). In the majority of RARS-T cases with SF3B1 and JAK2 mutations, SF3B1 mutant allele load was higher than JAK2 (89, 121), but this is not without conjecture (150). Moreover, SF3B1 mutations are rare in ET (0–3%) (63, 64), suggesting that many RARS-T cases evolve from RARS/RCMD-RS with acquisition of a JAK2 mutation (121).

Myeloproliferative neoplasms

MPNs are chronic, clonal HSC disorders and include polycythemia vera (PV), ET, and PMF. MPNs are associated with an increased risk of thrombosis, bleeding, and progressive bone marrow fibrosis, and importantly, a substantial proportion of patients transform to AML. The prognosis of secondary AML is dismal.

Spliceosome mutations were reported in MPNs (6–9%) (63, 152) and PMF (34%) (153). Mutations in SF3B1 (5–7%) (126, 132, 153), SRSF2 (3–17%) (132, 154, 155), U2AF1 (3–16%) (132, 153), and ZRSR (6%) (132) were detected in PMF patients. SF3B1 mutations most frequently involved K700E, K666T/N/M, H662D, and N626S codons (126). SF3B1 mutations were associated with marked splenomegaly and RS in PMF cases. There was no association between SF3B1 mutation and thrombosis and it did not influence survival (126). SRSF2 mutations were detected in 3–17% of PMF cases, mainly missense involving P95H/L/R/S and 24 bp deletion P95_R102del8 (126, 132, 154). Mutations in SRSF2, but not in other SF, are seen in high frequencies in patients with MPNs who later transform to AML (11.8% versus 2.8%) (132). SRSF2 mutations were associated with advanced age, IDH1/2 mutations, higher DIPSS-plus risk category, shortened OS, and LFS (154). The adverse effect of SRSF2 mutation on survival was independent of DIPSS-plus and IDH mutations (126). SRSF2 mutations were more frequent in AML transformed from MPN (19%) compared with AML progressed from MDS (4.8%) or *de novo* AML (5.6%) (132). Importantly, SRSF2 mutations were associated with poor OS in MPN patients who underwent AML transformation, independent of age and cytogenetic risk classification. Mutational analysis of paired samples from patients with chronic phase MPN who subsequently transformed to AML showed multiple genetic differences between chronic MPN state and AML state within individual

patients. Interestingly, mutations in SRSF2, when present in leukemic state, could also be identified in the earlier chronic MPN samples. This suggests that SRSF2 mutations are acquired early in disease pathogenesis and/or are required for leukemic transformation.

U2AF1 mutations are frequent spliceosome pathway mutations in PMF (3–16%) and commonly involve Q157 or S34 codons (132, 153). U2AF1 mutations are associated with anemia and thrombocytopenia and are detected mainly in DIPSS-plus intermediate-2 and high risk disease groups (155). They cluster with JAK2 (V617F), ASXL1 mutations and normal karyotype. They do not carry an independent prognostic relevance for either OS or LFS. ZRSR were detected in 5.6% of PMF and 1.8% of secondary AML from MPNs.

SRSF2 mutations are relatively common in PMF and independently predictive of poor outcome. SRSF2 mutations are more common in PMF patients who transform to AML and importantly are detected at similar frequencies during the chronic phase and at the time of transformation, raising the possibility that they might play a significant role in leukemic transformation. The clinical implication is that PMF patients harboring SRSF2 mutations might warrant more aggressive treatment such as with allogeneic stem cell transplantation in suitable patients.

Acute myeloid leukemia

The prevalence of spliceosome mutation was more frequent (31%) in MDS-AML compared to *de novo* AML (6.6%). SRSF2, SF3B1, and U2AF1 mutations were detected in 10%, 4%, and 0% of therapy related-myeloid neoplasm (t-MN) cases respectively (156). SRSF2 mutations were significantly higher in AML following genitourinary tumors compared with other t-MN.

Chronic lymphocytic leukemia

CLL, the most common leukemia of adults in Western countries, is a neoplasm of B lymphocytes. The clinical course of CLL ranges from a very indolent disorder with a normal lifespan, to a rapidly progressive disease leading to death. Identification of biological markers such as somatic hypermutation in the variable region of the immunoglobulin heavy chain (IGHV) gene, expression of ZAP and detection of cytogenetic abnormalities have improved our ability to predict a more aggressive disease course, but there is still significant variability within each risk group. Whole exome sequencing of CLL cases showed that >90% cases had at least one somatic mutation involving mRNA splicing, Toll-like receptor signaling, and apoptosis pathways among others (127). One of the most surprising findings arising from these studies was the discovery of SF3B1 mutation as a putative candidate driver gene in CLL. SF3B1 mutations, reported in 3.6–18.4% of CLL cases (81, 127, 128, 157–159), were among the most frequently identified somatic mutations, together with other putative driver mutations such as TP53, ATM, MYD88, BIRC3, and NOTCH1. The reported variability in SF3B1 mutation rates is mainly due to the patient population studied, for example, SF3B1 mutation rate was lower (1/63; 1.5%) in monoclonal B-cell lymphocytosis, a CLL precursor condition (160) and unselected, population-based Scandinavian CLL cohort (3.6%) (158), while SF3B1 mutation rates were higher (17%) in patients requiring frontline therapy for CLL (161) and fludarabine refractory patients (17%) (128, 162). As in the myeloid malignancies, the vast majority of SF3B1 mutations are localized to HEAT repeats 5-8 with K700E and K666 mutations predominantly in diagnosis CLL while K700E and H662 mutations were more common in fludarabine-refractory CLL. Intriguingly, a recurrent G742D mutation has been reported in several CLL studies, which is not seen in myeloid malignancies. To date, SF3B1 mutations have not been detected in Hodgkin's lymphoma, ALL or CML (87, 127, 163).

CLL patients with SF3B1 mutation have an advanced disease at diagnosis and adverse biological features, such as high β2-microglobulin and IGHV loci without mutations, compared to individuals with $SF3B1^{WT}$ (128, 157, 158). SF3B1 mutations occurred in tumors with del(11q), which are associated with poor prognosis CLL (81, 157). Furthermore, SF3B1 is associated with rapid disease progression, earlier time to treatment, shorter treatment free survival and poor OS (79, 81, 128, 157, 158, 160, 161, 164). In a study of 1160 CLL patients, $SF3B1^{Mut}$, IGHV mutational status and del(11q) were the only independent genetic markers for shorter time to treatment, whereas $SF3B1^{Mut}$, IGHV mutational status and TP53 disruption were associated with inferior OS (157). Similarly, recent analysis of the GCLLSG-CLL8 study showed that patients with SF3B1 and TP53 mutation had significantly lower PFS and a trend toward inferior OS following frontline fludarabine-based therapy (159). Taken together, these findings suggest that molecular markers such as TP53, SF3B1, and NOTCH1 mutation can potentially refine the current CLL prognostic risk stratification. Rossi and col-

leagues classified CLL cases (n = 1274) into four risk groups (high, intermediate, low, and very low risk groups) by integrating mutational and cytogenetic information. Patients harboring SF3B1 and/or NOTCH1 mutation and/or del(11q) are grouped as intermediate risk group with 10 years survival of 37% (165).

Fludarabine-based therapy has significantly improved PFS and OS of CLL patients and is a standard first line therapy; however, some patients are refractory. TP53, SF3B1, and NOTCH1 mutations were reported in 37.4%, 17.5%, and 13.4% of fludarabine refractory patients respectively (162). In these patients NOTCH1 and SF3B1 mutations were mutually exclusive. SF3B1 mutation did not adversely impact the OS and PFS following alemtuzumab therapy. Similarly, TP53, SF3B1, and NOTCH1 mutation status did not have significant impact on the OS and EFS in high risk CLL patients undergoing reduced-intensity allogeneic stem cell transplantation (166).

Intertumoral (79, 81) and intratumoral (167, 168) genetic heterogeneity is well-known in CLL. Analysis of the clonal and subclonal mutations in a large series of CLL cases including longitudinal case studies suggest that SF3B1 mutation is typically a subclonal genetic event in CLL, and hence is likely to be a non-initiating event but involved in disease progression (169). The presence of SF3B1 subclonal driver mutation is associated with poorer clinical outcome and was independent of confounding factors such as IGHV mutation status, identity of driver mutation and the presence of cytogenetic abnormalities.

At present, it remains unclear how SF3B1 mutation impacts CLL at a cellular level. Consistent with the critical role of SF3B1 in splicing, altered splicing has been demonstrated in CLL cases with SF3B1 mutations. Quesada *et al.* (79) compared splicing in CLL with and without mutated SF3B1 by exon arrays and RNA sequencing. They detected relatively few transcripts with altered splicing in CLL with mutated SF3B1, suggesting that the SF3B1 mutation does not exert a global effect on splicing, but rather only affects specific target transcripts (79). Quantitative PCR analysis showed enhanced expression of truncated mRNA from candidate splicing target genes in SF3B1-mutated cases. The new isoforms included truncated SLC23A2 and TCIRG1. In addition, one of the novel splicing sites affected FOXP1, encoding a forkhead TF whose altered expression has been linked to diffuse large B-cell non-Hodgkin's lymphoma. The expression of newly identified FOXP1 transcript was three times higher in SF3B1 mutated CLL cases than cases without mutation (127). Wang *et al.* (81) reported intron retention in known target genes of spliceosome inhibition. SF3B1 inhibitors alter splicing of a narrow spectrum of transcripts derived from genes involved in cancer-related processes, including cell cycle control, angiogenesis, and apoptosis. SF3B1 mutations led to mistakes in the splicing of these and other specific transcripts that affect the pathogenesis of CLL (81).

Opportunities for targeted therapies against mutated SF and AS variants in HMs

Splicing of pre-mRNA plays a crucial role in cell homeostasis; hence, the inhibition of mutant spliceosome using spliceosome modulators is a potentially effective and innovative therapeutic approach. As most of the SF mutations are heterozygous, it may be that homozygous mutation of spliceosomal gene is lethal. Cancer cells carrying spliceosome mutation already have compromised pre-mRNA splicing; hence, these cells may be more sensitive to the spliceosome modulators than non-mutated cells, a concept of synthetic lethality. One such example is MDS cells harboring del5q which are selectively sensitive to lenalidomide (170). The spliceosome modulators (pladienolides, FR901464 and E7107) are potent cytotoxic agents with IC_{50} values in the nM range against human cancer cell lines while being less toxic to normal cells in culture.

Pladienolides (A to G) are novel, unusual 12-membered macrolides isolated from *Streptomyces platensis* Mer-11107 (171, 172). Six of these seven pladienolides have an inhibitory activity against vascular endothelial growth factor (VEGF) expression and cancer cell proliferation *in vitro* (171, 172), pladienolide B being the most selective. Pladienolide B has been shown to arrest cell-cycle progression during the G1 phase and G2/M transition of the cell cycle *in vitro* (172) and also to inhibit tumor growth in several human cancer xenograft models. It impairs *in vivo* splicing in a dose-dependent manner by binding to SF3B3 (173). Its binding affinity with the SF3B complex is highly correlated with its inhibitory activities against reporter gene expression and cell proliferation suggesting that pladienolide exerts its potent activity by targeting SF3B3. Recent *in vitro* studies have shown that pladienolide-resistant cell lines have mutation R1074H in the SF3B1, which impairs the binding affinity of pladienolide suggesting that it may bind to an interface of SF3B3/SF3B1 (174). E7107, a derivative of pladienolide, demonstrated anti-tumor activity in human xenograft models

without significant toxicity (173, 175) and rapidly progressed to a Phase I clinical trials. However, this study was suspended; the reason remains unclear.

FR901463, FR901464, and FR901465, another group of natural spliceosome inhibitors, were isolated from fermentation broth of *Pseudomonas* species No. 2663 (176, 177). These compounds remarkably enhanced transcriptional activity of a Simian virus 40 (SV40) promoter, induced G1 and G2/M phase arrest and internucleosomal DNA fragmentation (176, 177). FR901464 exhibited anti-proliferative effect *in vitro*, inhibited growth of various tumors in murine model and prolonged the life of mice bearing the ascitic tumors (178). Meayamycin, a potent analogue of FR901464, inhibited *in vitro* proliferation of human cancer cell lines at picomolar concentration (178, 179). Spliceostatin A, a methylketal derivative of FR901464, also showed significant activity against various cancer cell lines (180). Sudemycins are analogues of FR901464 with a significantly better chemical stability (181). New synthetic spliceosome modulators are being developed that demonstrate the anti-tumor efficacy in animal studies (182).

A series of elegant experiments have shown that Spliceostatin A binds to the SF3B1/SF3B3 component of the SF3b complex (180). Spliceostatin A inhibited *in vitro* and *in vivo* splicing in a dose-dependent manner and caused the accumulation of pre-mRNA or partially spliced mRNA in the nucleus, an enlargement of the snRNP-enriched nuclear speckles and a reduction in cytoplasmic mRNA level by affecting the nuclear to cytoplasmic export of mRNA. Surprisingly, however, Spliceostatin A also allows nuclear export of unspliced mRNA and accumulation of proteins containing intron derived sequences. Culture with FR901464 and Spliceostatin A induced the accumulation of a C-terminal truncated form of p27 in multiple cancer cell lines while the expression of cyclin A and cyclin-dependent kinases remained unchanged. SF3b knockdown also led to similar changes suggesting that C-terminal truncated p27 and mislocalization of spliceosomes are the result of SF3b inhibition by Spliceostatin A (180). The C-terminal truncated p27 retains its activity as a cell cycle inhibitor, but is not degraded by the proteasome inhibitor due to the lack of C-terminus. This leads to the accumulation of C-terminal truncated p27 in tumor cells contributing toward their cell cycle arrest and the potential anti-cancer effects of SF3b modulation.

Another therapeutic approach is the modulation of AS to prevent the generation of oncogenic forms of some proteins such as genes encoding for VEGF and signal transducer and activator of transcription 3 (STAT3). Borrelidin is a potent inhibitor of angiogenesis which targets a spliceosome-associated protein WBP4 (FBP21), leading to modification of the ratio of VEGF isoforms in favor of anti-angiogenesis (183). STAT3 is constitutively activated in a number of cancers and has two isoforms, full length STAT3a and a shorter STAT3b. Splicing variant STAT3b uses an alternative 3′SS within exon 23 that leads to a truncated isoform lacking the C-terminal transactivation domain. STAT3b can act as a dominant negative regulator of transcription and promote apoptosis. An antisense oligonucleotide, targeting a splicing enhancer that regulates STAT3 exon 23 AS, specifically promotes a shift of expression from STAT3a to STAT3b causing cell cycle arrest and apoptosis in cell lines and tumor regression in a cancer xenograft model (184).

Spliceosome inhibitor studies have been focused on solid tumor cell lines and murine models with few human studies. There are many unanswered questions. What is the effect of these spliceosome modulators in patient cells harboring different spliceosome mutations? What are the critical proteins/pathways/networks impacted by SF mutations? Are all mutant SF equally susceptible to inhibition? What SF mutations are amenable to synthetic lethality strategies? Do any SF mutations display oncogene addiction? Can drugs be used to restore normal splicing?

Concluding remarks

Splicing is a highly complex and intricately regulated process enabling cell-, signal-, and temporal-specific cues or requirements to be enacted within the cell. We are beginning to understand more about the importance of differential splicing patterns during normal hemopoiesis of HSC activation with subsequent cascading differentiation programs and rapidly expanding cell populations, and how perturbations to SF and splicing mechanisms can impact on this to cause a range of HM. This knowledge is beginning to be incorporated into diagnostic testing and prognostic predictions. It has also opened avenues to targeted therapeutic approaches, some of which are entering preclinical trials and even the clinic. As we learn more, novel strategies will become available to tackle these often difficult to treat diseases.

References

1. Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 2012;**22**:1760–1774.
2. Jang JH. Identification and characterization of soluble isoform of fibroblast growth factor receptor 3 in human SaOS-2 osteosarcoma cells. Biochem Biophys Res Commun 2002;**292**:378–382.
3. Coutts JC, Gallagher JT. Receptors for fibroblast growth factors. Immunol Cell Biol 1995;**73**:584–589.
4. Schuster K, Fan L, Harris LC. MDM2 splice variants predominantly localize to the nucleoplasm mediated by a COOH-terminal nuclear localization signal. Mol Cancer Res 2007;**5**:403–412.
5. Liu F, Gong CX. Tau exon 10 alternative splicing and tauopathies. Mol Neurodegener 2008;**3**:8.
6. Neueder A, Achilli F, Moussaoui S, Bates GP. Novel Isoforms of Heat Shock Transcription Factor 1, HSF1gammaalpha and HSF1gammabeta, Regulate Chaperone Protein Gene Transcription. J Biol Chem 2014;**289**:19894–19906.
7. Will CL, Luhrmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol 2011;**3**:1–23.
8. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 2003;**12**:5–14.
9. Hegele A, et al. Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell 2012;**45**:567–580.
10. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell 2009;**136**:701–718.
11. Wang ET, et al. Alternative isoform regulation in human tissue transcriptomes. Nature 2008;**456**:470–476.
12. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 2008;**40**:1413–1415.
13. Turunen JJ, Niemela EH, Verma B, Frilander MJ. The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA 2013;**4**:61–76.
14. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 2004;**5**:89–99.
15. Savage KI, et al. Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. Mol Cell 2014;**54**:445–459.
16. Adamia S, et al. A genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets. Clin Cancer Res 2013;**20**:1135–1145.
17. Liu Q, Zhao S, Su PF, Yu S. Gene and isoform expression signatures associated with tumor stage in kidney renal clear cell carcinoma. BMC Syst Biol 2013;**7**(Suppl 5):S7.
18. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene 2013; doi: 10.1038/onc.2013.533. [Epub ahead of print].
19. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev 2010;**24**:2343–2364.
20. Eswaran J, et al. RNA sequencing of cancer reveals novel splicing alterations. Sci Rep 2013;**3**:1689.
21. Misquitta-Ali CM, et al. Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. Mol Cell Biol 2011;**31**:138–150.
22. Pesson M, Eymin B, De La Grange P, Simon B, Corcos L. A dedicated microarray for in-depth analysis of pre-mRNA splicing events: application to the study of genes involved in the response to targeted anticancer therapies. Mol Cancer 2014;**13**:9.
23. Pesson M, et al. A gene expression and pre-mRNA splicing signature that marks the adenoma-adenocarcinoma progression in colorectal cancer. PLoS ONE 2014;**9**:e87761.
24. Surget S, Khoury MP, Bourdon J. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther 2014;**7**:57–68.
25. Saito S, et al. CD44v6 expression is related to mesenchymal phenotype and poor prognosis in patients with colorectal cancer. Oncol Rep 2013;**29**:1570–1578.
26. Shi J, Zhou Z, Di W, Li N. Correlation of CD44v6 expression with ovarian cancer progression and recurrence. BMC Cancer 2013;**13**:182.
27. Cha JY, Lambert QT, Reuther GW, Der CJ. Involvement of fibroblast growth factor receptor 2 isoform switching in mammary oncogenesis. Mol Cancer Res 2008;**6**:435–445.
28. Harper SJ, Bates DO. VEGF-A splicing: the key to anti-angiogenic therapeutics? Nat Rev Cancer 2008;**8**:880–887.
29. Zong FY, et al. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. PLoS Genet 2014;**10**:e1004289.
30. Kunnimalaiyaan M, Chen H. Tumor suppressor role of Notch-1 signaling in neuroendocrine tumors. Oncologist 2007;**12**:535–542.
31. Lobry C, Oh P, Aifantis I. Oncogenic and tumor suppressor functions of Notch in cancer: it's NOTCH what you think. J Exp Med 2011;**208**:1931–1935.
32. Weng AP, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science 2004;**306**:269–271.
33. Gianfelici V. Activation of the NOTCH1 pathway in chronic lymphocytic leukemia. Haematologica 2012;**97**:328–330.
34. Avila JL, Kissil JL. Notch signaling in pancreatic cancer: oncogene or tumor suppressor? Trends Mol Med 2013;**19**:320–327.
35. Pritsker M, Doniger TT, Kramer LC, Westcot SE, Lemischka IR. Diversification of stem cell molecular repertoire by alternative splicing. Proc Natl Acad Sci USA 2005;**102**:14290–14295.
36. Tanaka T, et al. An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. EMBO J 1995;**14**:341–350.
37. Popovic R, Erfurth F, Zeleznik-Le N. Transcriptional complexity of the HOXA9 locus. Blood Cells Mol Dis 2008;**40**:156–159.
38. Nam DK, Honoki K, Yu M, Yunis JJ. Alternative RNA splicing of the MLL gene in normal and malignant cells. Gene 1996;**178**:169–175.
39. Golub TR, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;**286**:531–537.
40. Studler CR, et al. The leukemogenicity of Hoxa9 depends on alternative splicing. Leukemia 2014;**28**:1838–1843.
41. Kastner P, Chan S. PU.1: a crucial and versatile player in hematopoiesis and leukemia. Int J Biochem Cell Biol 2008;**40**:22–27.
42. Zhang P, et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc Natl Acad Sci USA 1999;**96**:8705–8710.
43. Hallier M, Lerga A, Barnache S, Tavitian A, Moreau-Gachelin F. The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. J Biol Chem 1996;**273**:4838–4842.
44. Hallier M, Tavitian A, Moreau-Gachelin F. The transcription factor Spi-1/PU.1 binds RNA and interferes with the RNA-binding protein p54nrb. J Biol Chem 1998;**271**:11177–11181.
45. Delva L, Gallais I, Guillouf C, Denis N, Orvain C, Moreau-Gachelin F. Multiple functional domains of the oncoproteins Spi-1/PU.1 and TLS are involved in their opposite splicing effects in erythroleukemic cells. Oncogene 2004;**23**:4389–4399.
46. Guillouf C, Gallais I, Moreau-Gachelin F. Spi-1/PU.1 oncoprotein affects splicing decisions in a promoter binding-dependent manner. J Biol Chem 2006;**281**:19145–19155.
47. Zhou J, et al. Enhanced activation of STAT pathways and overexpression of survivin confer resistance to FLT3 inhibitors and could be therapeutic targets in AML. Blood 2009;**113**:4052–4062.
48. Altieri DC. Survivin, cancer networks and pathway-directed drug discovery. Nat Rev Cancer 2008;**8**:61–70.
49. Wagner M, Schmelz K, Wuchter C, Ludwig WD, Dorken B, Tamm I. In vivo expression of survivin and its splice variant survivin-2B: impact on clinical outcome in acute myeloid leukemia. Int J Cancer 2006;**119**:1291–1297.
50. Carter BZ, et al. Survivin is highly expressed in CD34(+)38(-) leukemic stem/progenitor cells and predicts poor clinical outcomes in AML. Blood 2012;**120**:173–180.
51. Fukuda S, et al. Survivin mediates aberrant hematopoietic progenitor cell proliferation and acute leukemia in mice induced by internal tandem duplication of Flt3. Blood 2009;**114**:394–403.

52. Moore AS, et al. BIRC5 (survivin) splice variant expression correlates with refractory disease and poor outcome in pediatric acute myeloid leukemia: a report from the Children's Oncology Group. Pediatr Blood Cancer 2014;**61**:647–652.
53. Kramarzova K, et al. Real-time PCR quantification of major Wilms' tumor gene 1 (WT1) isoforms in acute myeloid leukemia, their characteristic expression patterns and possible functional consequences. Leukemia 2012;**26**:2086–2095.
54. Scott LM, Rebel VI. Acquired mutations that affect pre-mRNA splicing in hematologic malignancies and solid tumors. J Natl Cancer Inst 2013;**105**:1540–1549.
55. Risso G, et al. Modification of Akt by SUMO conjugation regulates alternative splicing and cell cycle. Cell Cycle 2013;**12**:3165–3174.
56. Ng KP, et al. A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. Nat Med 2012;**18**:521–528.
57. Anczukow O, et al. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. Nat Struct Mol Biol 2012;**19**:220–228.
58. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;**144**:646–674.
59. Arana-Trejo RM, et al. BCR/ABL p210, p190 and p230 fusion genes in 250 Mexican patients with chronic myeloid leukaemia (CML). Clin Lab Haematol 2002;**24**:145–150.
60. Anensen N, Oyan AM, Bourdon JC, Kalland KH, Bruserud O, Gjertsen BT. A distinct p53 protein isoform signature reflects the onset of induction chemotherapy for acute myeloid leukemia. Clin Cancer Res 2006;**12**:3985–3992.
61. Konopleva M, et al. Expression and function of leptin receptor isoforms in myeloid leukemia and myelodysplastic syndromes: proliferative and anti-apoptotic activities. Blood 1999;**93**:1668–1676.
62. Walter MJ, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia 2013;**27**:1275–1282.
63. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;**478**:64–69.
64. Papaemmanuil E, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 2011;**365**:1384–1395.
65. Puig O, Bragado-Nilsson E, Koski T, Seraphin B. The U1 snRNP-associated factor Luc7p affects 5′ splice site selection in yeast and human. Nucleic Acids Res 2007;**35**:5874–5885.
66. Busque L, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat Genet 2012;**44**:1179–1181.
67. Chan SM, Majeti R. Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. Int J Hematol 2013;**98**:648–657.
68. Papaemmanuil E, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;**122**:3616–3627; quiz 3699.
69. Walter MJ, et al. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med 2012;**366**:1090–1098.
70. Lin CC, et al. SF3B1 mutations in patients with myelodysplastic syndromes: the mutation is stable during disease evolution. Am J Hematol 2014;**89**:E109–E115.
71. Wu SJ, et al. Clinical implications of U2AF1 mutation in patients with myelodysplastic syndrome and its stability during disease progression. Am J Hematol 2013;**88**:E277–E282.
72. Keightley MC, et al. In vivo mutation of pre-mRNA processing factor 8 (Prpf8) affects transcript splicing, cell survival and myeloid differentiation. FEBS Lett 2013;**587**:2150–2157.
73. Mian SA, et al. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica 2013;**98**:1058–1066.
74. Das BK, Xia L, Palandjian L, Gozani O, Chyung Y, Reed R. Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. Mol Cell Biol 1999;**19**:6796–6802.
75. Wang C, et al. Depletion of Sf3b1 impairs proliferative capacity of hematopoietic stem cells but is not sufficient to induce myelodysplasia. Blood 2014;**123**:3336–3343.
76. Matsunawa M, et al. Haploinsufficiency of Sf3b1 leads to compromised stem cell function but not to myelodysplasia. Leukemia 2014;**28**:1844–1850.
77. Golas MM, Sander B, Will CL, Luhrmann R, Stark H. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. Mol Cell 2005;**17**:869–883.
78. Hahn CN, Scott HS. Spliceosome mutations in hematopoietic malignancies. Nat Genet 2011;**44**:9–10.
79. Quesada V, Ramsay AJ, Lopez-Otin C. Chronic lymphocytic leukemia with SF3B1 mutation. N Engl J Med 2012;**366**:2530.
80. Visconte V, et al. SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. Leukemia 2012a;**26**:542–545.
81. Wang L, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 2011;**365**:2497–2506.
82. Biankin AV, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;**491**:399–405.
83. Network CGA. Comprehensive molecular portraits of human breast tumours. Nature 2012;**490**:61–70.
84. Furney SJ, et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. Cancer Discov 2013;**3**:1122–1129.
85. Harbour JW, Roberson ED, Anbunathan H, Onken MD, Worley LA, Bowcock AM. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nat Genet 2013;**45**:133–135.
86. Kong Y, Krauthammer M, Halaban R. Rare SF3B1 R625 mutations in cutaneous melanoma. Melanoma Res 2014;**24**:332–334.
87. Rozovski U, Keating M, Estrov Z. The significance of spliceosome mutations in chronic lymphocytic leukemia. Leuk Lymphoma 2012;**54**:1364–1366.
88. Damm F, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood 2012;**119**:3211–3218.
89. Malcovati L, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood 2011;**118**:6239–6246.
90. Visconte V, et al. SF3B1 haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. Blood 2012b;**120**:3173–3186.
91. Visconte V, et al. Distinct iron architecture in SF3B1-mutant myelodysplastic syndrome patients is linked to an SLC25A37 splice variant with a retained intron. Leukemia 2014; doi: 10.1038/leu.2014.170. [Epub ahead of print].
92. Makishima H, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood 2012;**119**:3203–3210.
93. Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature 1999;**402**:832–835.
94. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. Nat Struct Mol Biol 2004;**11**:257–264.
95. Graubert TA, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet 2011;**44**:53–57.
96. Brooks AN, et al. A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PLoS ONE 2014;**9**:e87361.
97. Przychodzen B, et al. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood 2013;**122**:999–1006.
98. Ilagan JO, et al. U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res 2014; doi: 10.1101/gr.181016.114. [Epub ahead of print].
99. Patnaik MM, et al. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. Blood 2012;**119**:569–572.
100. Phelan MM, Goult BT, Clayton JC, Hautbergue GM, Wilson SA, Lian LY. The structure and selectivity of the SR protein SRSF2 RRM domain with RNA. Nucleic Acids Res 2012;**40**:3232–3244.
101. Grainger RJ, Beggs JD. Prp8 protein: at the heart of the spliceosome. RNA 2005;**11**:533–557.
102. Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 2013;**493**:638–643.
103. Konforti BB, Konarska MM. U4/U5/U6 snRNP recognizes the 5′ splice site in the absence of U2 snRNP. Genes Dev 1994;**8**:1962–1973.

104. Kurtovic-Kozaric A, et al. PRPF8 defects cause missplicing in myeloid malignancies. Leukemia 2014; doi: 10.1038/leu.2014.144. [Epub ahead of print].
105. Isono K, Mizutani-Koseki Y, Komori T, Schmidt-Zachmann MS, Koseki H. Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. Genes Dev 2005;**19**:536–541.
106. Xiao R, et al. Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. Mol Cell Biol 2007;**27**:5393–5402.
107. Edmond V, et al. Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. EMBO J 2011;**30**:510–523.
108. Edmond V, Brambilla C, Brambilla E, Gazzeri S, Eymin B. SRSF2 is required for sodium butyrate-mediated p21(WAF1) induction and premature senescence in human lung carcinoma cell lines. Cell Cycle 2011;**10**:1968–1977.
109. Thol F, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood 2012;**119**:3578–3584.
110. Mo S, Ji X, Fu XD. Unique role of SRSF2 in transcription activation and diverse functions of the SR and hnRNP proteins in gene expression regulation. Transcription 2013;**4**:251–259.
111. Bejar R, et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J Clin Oncol 2012;**30**:3376–3382.
112. Xu L, et al. Genomic landscape of CD34+ hematopoietic cells in myelodysplastic syndrome and gene mutation profiles as prognostic markers. Proc Natl Acad Sci USA 2014;**111**:8589–8594.
113. Greif PA, et al. GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. Blood 2012;**120**:395–403.
114. Ma X, Does M, Raza A, Mayne ST. Myelodysplastic syndromes: incidence and survival in the United States. Cancer 2007;**109**:1536–1542.
115. Vardiman JW, Bennett JM, Bain BJ, Baumann I, Thiele J, Orazi A. *Myelodysplastic/Myeloproliferative Neoplasms, Unclassifiable*. Lyon, France: IARC, 2008.
116. Vardiman JW, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 2009;**114**:937–951.
117. Greenberg P, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 1997;**89**:2079–2088.
118. Greenberg PL, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood 2012;**120**:2454–2465.
119. Haase D, et al. New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. Blood 2007;**110**:4385–4395.
120. Abdel-Wahab O, Figueroa ME. Interpreting new molecular genetics in myelodysplastic syndromes. Hematology 2012;**2012**:56–64.
121. Haferlach T, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 2014;**28**:241–247.
122. Bejar R, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med 2011;**364**:2496–2506.
123. Malcovati L, et al. Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia. Blood 2014;**124**:1513–1521.
124. Visconte V, et al. SF3B1 mutations are infrequently found in non-myelodysplastic bone marrow failure syndromes and mast cell diseases but, if present, are associated with the ring sideroblast phenotype. Haematologica 2013;**98**:e105–e107.
125. Damm F, et al. SF3B1 mutations in myelodysplastic syndromes: clinical associations and prognostic implications. Leukemia 2012;**26**:1137–1140.
126. Lasho TL, et al. SF3B1 mutations in primary myelofibrosis: clinical, histopathology and genetic correlates among 155 patients. Leukemia 2012a;**26**:1135–1137.
127. Quesada V, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 2012;**44**:47–52.
128. Rossi D, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood 2011;**118**:6904–6908.
129. Graubert TA, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet 2012;**44**:53–57.
130. Kambach C, Walke S, Nagai K. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. Curr Opin Struct Biol 1999;**9**:222–230.
131. Itzykson R, et al. Prognostic score including gene mutations in chronic myelomonocytic leukemia. J Clin Oncol 2013;**31**:2428–2436.
132. Zhang SJ, et al. Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood 2012;**119**:4480–4485.
133. Pellagatti A, et al. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood 2006;**108**:337–345.
134. Boultwood J, et al. The role of the iron transporter ABCB7 in refractory anemia with ring sideroblasts. PLoS ONE 2008;**3**:e1970.
135. Nilpour M, et al. Gene expression profiling of erythroblasts from refractory anaemia with ring sideroblasts (RARS) and effects of G-CSF. Br J Haematol 2010;**149**:844–854.
136. Nilpour M, et al. The transporter ABCB7 is a mediator of the phenotype of acquired refractory anemia with ring sideroblasts. Leukemia 2013;**27**:889–896.
137. Bacher U, Haferlach T, Schnittger S, Kreipe H, Kroger N. Recent advances in diagnosis, molecular pathology and therapy of chronic myelomonocytic leukaemia. Br J Haematol 2011;**153**:149–167.
138. Germing U, Kundgen A, Gattermann N. Risk assessment in chronic myelomonocytic leukemia (CMML). Leuk Lymphoma 2004;**45**:1311–1318.
139. Onida F, et al. Prognostic factors and scoring systems in chronic myelomonocytic leukemia: a retrospective analysis of 213 patients. Blood 2002;**99**:840–849.
140. Kantarjian H, et al. Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. Cancer 2008;**113**:1351–1361.
141. Worsley A, et al. Prognostic features of chronic myelomonocytic leukaemia: a modified Bournemouth score gives the best prediction of survival. Br J Haematol 1988;**68**:17–21.
142. Kar SA, et al. Spliceosomal gene mutations are frequent events in the diverse mutational spectrum of chronic myelomonocytic leukemia but largely absent in juvenile myelomonocytic leukemia. Haematologica 2013;**98**:107–113.
143. Patnaik MM, et al. Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. Am J Hematol 2013a;**88**:201–206.
144. Meggendorfer M, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). Blood 2012;**120**:3080–3088.
145. Patnaik MM, et al. ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two-center study of 466 patients. Leukemia 2014;10.1038/leu.2014.125. [Epub ahead of print].
146. Patnaik MM, et al. Mayo prognostic model for WHO-defined chronic myelomonocytic leukemia: ASXL1 and spliceosome component mutations and outcomes. Leukemia 2013b;**27**:1504–1510.
147. Hirabayashi S, et al. Spliceosomal gene aberrations are rare, coexist with oncogenic mutations, and are unlikely to exert a driver effect in childhood MDS and JMML. Blood 2012;**119**:e96–e99.
148. Gelsi-Boyer V, et al. Molecular similarity between myelodysplastic form of chronic myelomonocytic leukemia and refractory anemia with ring sideroblasts. Haematologica 2013;**98**:576–583.
149. Ceesay MM, et al. The JAK2 V617F mutation is rare in RARS but common in RARS-T. Leukemia 2006;**20**:2060–2061.
150. Jeromin S, et al. High frequencies of SF3B1 and JAK2 mutations in refractory anemia with ring sideroblasts associated with marked thrombocytosis strengthen the assignment to the category of myelodysplastic/myeloproliferative neoplasms. Haematologica 2013;**98**:e15–e17.
151. Broseus J, et al. Age, JAK2(V617F) and SF3B1 mutations are the main predicting factors for survival in refractory anaemia with ring

sideroblasts and marked thrombocytosis. Leukemia 2013;**27**:1826–1831.
152. Nangalia J, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. N Engl J Med 2013;**369**:2391–2405.
153. Tefferi A, et al. U2AF1 mutations in primary myelofibrosis are strongly associated with anemia and thrombocytopenia despite clustering with JAK2V617F and normal karyotype. Leukemia 2014a;**28**:431–433.
154. Lasho TL, et al. SRSF2 mutations in primary myelofibrosis: significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival. Blood 2012b;**120**:4168–4171.
155. Tefferi A, et al. CALR vs JAK2 vs MPL-mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons. Leukemia 2014b;**28**:1472–1477.
156. Voso MT, et al. Mutations of epigenetic regulators and of the spliceosome machinery in therapy-related myeloid neoplasms and in acute leukemias evolved from chronic myeloproliferative diseases. Leukemia 2013;**27**:982–985.
157. Jeromin S, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. Leukemia 2014;**28**:108–117.
158. Mansouri L, et al. NOTCH1 and SF3B1 mutations can be added to the hierarchical prognostic classification in chronic lymphocytic leukemia. Leukemia 2013;**27**:512–514.
159. Stilgenbauer S, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. Blood 2014;**123**:3247–3254.
160. Greco M, et al. Analysis of SF3B1 mutations in monoclonal B-cell lymphocytosis. Hematol Oncol 2013;**31**:54–55.
161. Oscier DG, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. Blood 2013;**121**:468–475.
162. Schnaiter A, et al. NOTCH1, SF3B1, and TP53 mutations in fludarabine-refractory CLL patients treated with alemtuzumab: results from the CLL2H trial of the GCLLSG. Blood 2013;**122**:1266–1270.
163. Yang J, et al. SF3B1 mutation is a rare event in Chinese patients with acute and chronic myeloid leukemia. Clin Biochem 2013;**46**:701–703.
164. Baliakas P, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia 2014; doi:10.1038/leu.2014.196. [Epub ahead of print].
165. Rossi D, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. Blood 2013;**121**:1403–1412.
166. Dreger P, et al. TP53, SF3B1, and NOTCH1 mutations and outcome of allotransplantation for chronic lymphocytic leukemia: six-year follow-up of the GCLLSG CLL3X trial. Blood 2013;**121**:3284–3288.
167. Schuh A, et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood 2012;**120**:4191–4196.
168. Stilgenbauer S, et al. Clonal evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival. Haematologica 2007;**92**:1242–1245.
169. Landau DA, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 2013;**152**:714–726.
170. List A, et al. Lenalidomide in the myelodysplastic syndrome with chromosome 5q deletion. N Engl J Med 2006;**355**:1456–1465.
171. Sakai T, Asai N, Okuda A, Kawamura N, Mizui Y. Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. II. Physico-chemical properties and structure elucidation. J Antibiot 2004a;**57**:180–187.
172. Sakai T, Sameshima T, Matsufuji M, Kawamura N, Dobashi K, Mizui Y. Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. I. Taxonomy, fermentation, isolation and screening. J Antibiot 2004b;**57**:173–179.
173. Kotake Y, et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol 2007;**3**:570–575.
174. Yokoi A, et al. Biological validation that SF3b is a target of the antitumor macrolide pladienolide. FEBS J 2011;**278**:4870–4880.
175. Folco EG, Coil KE, Reed R. The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. Genes Dev 2011;**25**:440–444.
176. Nakajima H, et al. New antitumor substances, FR901463, FR901464 and FR901465. II. Activities against experimental tumors in mice and mechanism of action. J Antibiot 1996;**49**:1204–1211.
177. Nakajima H, Sato B, Fujita T, Takase S, Terano H, Okuhara M. New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. J Antibiot 1996;**49**:1196–1203.
178. Albert BJ, Sivaramakrishnan A, Naka T, Czaicki NL, Koide K. Total syntheses, fragmentation studies, and antitumor/antiproliferative activities of FR901464 and its low picomolar analogue. J Am Chem Soc 2007;**129**:2648–2659.
179. Osman S, Albert BJ, Wang Y, Li M, Czaicki NL, Koide K. Structural requirements for the antiproliferative activity of pre-mRNA splicing inhibitor FR901464. Chemistry 2011;**17**:895–904.
180. Kaida D, et al. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol 2007;**3**:576–583.
181. Fan L, Lagisetti C, Edwards CC, Webb TR, Potter PM. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. ACS Chem Biol 2011;**6**:582–589.
182. Lagisetti C, et al. Synthetic mRNA splicing modulator compounds with in vivo antitumor activity. J Med Chem 2009;**52**:6979–6990.
183. Woolard J, et al. Borrelidin modulates the alternative splicing of VEGF in favour of anti-angiogenic isoforms. Chem Sci 2011;**2011**:273–278.
184. Zammarchi F, et al. Antitumorigenic potential of STAT3 alternative splicing modulation. Proc Natl Acad Sci USA 2011;**108**:17779–17784.

Appendix II. Characterisation of a compound in-cis GATA2 germline mutation in a pedigree presenting with myelodysplastic syndrome/acute myeloid leukaemia with concurrent thrombocytopaenia

Hahn, C.N., Brautigan, P.J., Chong, C.E., Janssan, A., Venugopal, P., Lee, Y., Tims, A.E., Horwitz, M.S., Klingler-Hoffmann, M. and Scott, H.S. (2015). Characterisation of a compound in-cis GATA2 germline mutation in a pedigree presenting with myelodysplastic syndrome/acute myeloid leukemia with concurrent thrombocytopenia. *Leukemia*, 29(8), 1795-97.

Appendix III: Summary of germline and somatic *GATA2* mutations and associated disease phenotypes

GATA2 Mutation	Overall Phenotypes	Number of patients	Publications
5'UTR	MDS, ID, AA	7	(66, 70, 157)
P41A	MDS	3	(158)
E44*	L	2	(159)
S54*	MDS, AML	1	(61)
R69Lfs*115	AML	1	Personal communication
V70Lfs*114	MDS	1	(61)
S71F	Malignant Melanoma	1	COSMIC
R78Pfs*107	MDS, AML, L	3	(36)
G81fs*	MDS, smouldering myeloma	1	(160, 161)
G82Rfs*103	MDS, ID	2	(66, 157, 162)
C85fs*	Autoimmunity	1	(163)
G101Afs*16	MDS, AML, ID	2	(66, 157, 160, 164, 165)
A103fs*116	MDS	3	(61)
L105Pfs*14	L, MDS, AML, ID	7	(36, 158)
S106fs*	ID, L	2	(61)
P125T	Lung Cancer	1	COSMIC
S129fs *	AML	1	(166)
S139Cfs*45	MDS	1	(61)
V140Cfs*44	MDS, ID	2	(157, 160, 161, 164)
A194Sfs*8	AML, L	1	(36)
S197fs*	MDS	1	(86)
G200fs*1	MDS, ID	5	(61, 163, 167)
S201*	AML	1	(168)
S201fs*	MDS	1	(169)

A203P	AML	1	(165)
R204*	AML, ID	1	(65)
V211Rfs*72	MDS	1	(61)
E224*	ID	1	(65)
L229Cfs*5	MDS	1	(61)
P245fs*	ID	1	(163)
P254L	MDS, ID	2	(38, 66)
D259fs*	MDS, ID	2	(66)
Y260fs*24	MDS	1	(160)
F265fs*	MDS	1	(157)
G268*	MDS, L	4	(168)
M1del290	L, MDS, ID	3	(66, 157, 160, 161)
S290*	MDS	1	(86)
N297D	AML	1	(170)
N297S	AML	1	(171)
C298Lfs*86	MDS, AML, ID	4	(168)
G299R	AML	1	(170)
T303S	AML	1	(170)
P304H	MDS, AML	2	(72, 170)
P304L	AML	1	(172)
L305V	MDS	1	Personal communication
W306fs*77	MDS	1	(173)
R307L	AML	1	(171)
R307Q	AML	1	(165)
R307W	AML, ETP ALL	1	(170, 174)
R308P	AML	5	(165, 172, 175-177)
R308Q	AML	2	(170, 178)
D309V	AML	1	(172)
T311Rfs*71	MDS, AML	1	(61)
N317D	AML	1	(165)

N317fs*	MDS, ID	2	(38, 66)
N317H	AML	2	(170, 172)
N317I	AML	3	(165, 171, 175, 176)
N317S	AML	2	(170, 175, 176)
A318_C319insS	MDS, AML	1	(72)
A318D	AML	2	(170, 171)
A318fs*12	MDS	2	(157, 160, 161)
A318G	AML	4	(170, 172, 179, 180)
A318T	AML	9	(165, 170-172, 179, 180)
A318V	AML	12	(58, 165, 168, 170, 175, 176) (172)
G320A	AML	1	(172)
G320D	AML	13	(170, 171, 175, 176, 178-180) (172)
G320V	AML	2	(170) (172)
L321F	AML	15	(165, 171, 175, 176, 179, 180) (172)
L321H*	AML	2	(170)
L321P	AML	3	(179, 180)
L321R	AML	2	(171, 175, 176)
L321V	AML	1	(61, 175, 176, 179, 180)
Y322C	AML	1	(170)
H323Qfs*61	MDS	1	(61)
K324E	AML	1	(172)
G346S	ID	1	Personal communication
Q328P	MDS, AML	1	(179, 180)
R330*	MDS, AML, L, ID	9	(65, 157, 160, 161, 181)
R330L	AML	7	(165, 170) (172) (182)
R330P	AML	1	(165)

R330Q	MDS, AML	5	(165, 171, 175, 176, 179, 180)
L332Qfs*60	MDS	1	(61)
L332Tfs*53	MDS, AML, L	2	(37, 38, 62)
R337*	MDS, AML, L, ID	4	(36, 61, 66, 157, 160, 161)
del 340-381	MDS, ID	3	(161, 163)
S340-N381del	MDS, ID	4	(38, 66, 157, 160)
A341_G346del	CML-BC	2	(180) Personal communication
A341fs*	L, ID	1	(163)
A341Pfs*45	MDS, AML, L, HL	1	(36)
A341Rfs*38	MDS, AML, L, ID, HL	1	(36)
A342Gfs*41	ID, AA	1	(164)
A342T	AML	2	(166)
R343_T347del	AML	1	(170)
R344Kfs*37	MDS	1	(61)
T347Rfs*38	MDS	1	(61)
C348_C351dup	AML	1	(170)
C348F	MDS	1	(61)
A350_N351ins8	AML	1	(177)
A350_C351dup	AML	1	(170)
N351D	AML	1	(171)
C352G	MDS, AML	1	(61)
T354K	AML	3	(171, 175, 176)
T354M	MDS, AML, ID	63	(37, 38, 63, 157, 158, 160, 162-164, 167, 183-185)
355delT	MDS, AML	2	(19)
T355_T356insT	AML	1	(171)
T356_N365del	MDS	1	(61)
T357A	MDS	2	(61)

T357S	AML	1	(166)
L359V	CML-BC	9	(62, 171)
R361C	MDS, AML, L, ID	6	(61, 66, 157, 163, 186)
R361del4	MDS	2	(157, 160, 161)
R361H	MDS, AML, ID	4	(61, 157, 161, 175, 176)
R361L	L, MDS, AML, ID, HL	1	(36)
R362*	MDS	2	(61, 170)
R362_N365del	ID	2	(38, 162, 164)
R362fs*24	MDS, AML	1	(58)
R362G	MDS, AML	4	(58, 170, 171)
R362P	MDS, AML	3	(58, 166)
R362Q	MDS, AML	17	(58, 72, 166, 170, 171, 175, 176)
G366R	AML	1	(171)
D367fs*15	MDS, ID	2	(157, 160, 161)
C370W	MDS	1	(61)
N371K	MDS, AML, ID	4	(38, 61, 157, 160, 162)
A372T	AML, ID	5	(65, 161, 163, 170, 171, 173)
C373del5	MDS, AML, L, ID	4	(157, 160, 161, 187)
C373R	MDS, AML, L, ID	1	(36)
L375F	ID	2	(157, 161)
L375I	AML	1	(166)
L375Pfs*12	MDS	1	(61)
Y376*	MDS	1	(61)
R384G	MDS	1	(173)
R384K	MDS	1	(173)
P385L	AML	1	(165)
M388_K389del	MDS	1	(86)

M388T	ID	4	(66, 157, 160)
M388V	ID	2	(65)
K390delK	MDS	3	(163, 173)
K390E	MDS	1	(61, 158)
R396L	AML	1	(171)
R396Q	MDS, AML, L, ID	21	(38, 65, 157, 158, 160, 162, 188, 189)
R396W	MDS, AML, ID	13	(38, 61, 66, 157, 160-162, 164, 190) Personal communication
R398Q	ID	5	(163)
R398W	MDS, AML, ID	17	(38, 64, 66, 157, 160, 162, 163, 188)
E415K	MDS, AML	1	(61)
F427fs*	MDS	1	(169)
L443I	MDS	1	(169)
S447R	MDS	2	(61)
S464I	AML	1	(171)
S473P	AML	1	(178)
G136Rfs*49	MDS, ID	1	(191)
G199fs*21	MDS, L, ID	2	(66, 157, 160, 163)
G200Vfs*18	MDS	2	(61)
del28 Intron 4 (between exon 4 & 5)	MDS, AML, ID, L	6	(66, 157, 164, 192)
Intron 4 (between exon 4 & 5)	MDS, AML, ID, L, AA	32	(37, 61, 66, 70, 157, 160, 161)
L321F R330Q #	AML	1	(165)

L321H L321F #	AML	1	(171)
L321H L379Q #	MDS, AML	1	(175, 176)
A318S T354M #	AML	1	(165)
A318V R293Q #	MDS, AML	1	(179, 180)
H26P G28fs*52 #	ID	7	(193)
N297S N317S #	MDS	1	(173)
N317H A318T #	MDS, AML	1	(179, 180)
R307L L321F #	MDS, AML	1	(175)
R307W G320V #	AML	1	(165)
R330* L321R #	JMML	1	(194)
R362dup S449S #	AML	1	(165)
L375V 355 bp ins #	AML	1	(156)
K390E G310fs* T117fs* #	AML	1	(86)
A372V, Chr3 del (1.39 copies) #	AML		(171)

Chr3 del (1.26 copies)	AML	1	(171)
Chr3 del (1.17 copies)	AML	1	(171)
Chr3 del (1.15 copies)	AML	1	(171)
Chr3 del (1.11 copies)	AML	2	(171)
Chr3 del (1.09 copies)	AML	1	(171)
Chr3 del (1.05 copies)	AML	1	(171)
Chr3 del (0.87 copies)	AML	1	(171)

MDS – Myelodysplastic Syndrome

AML – Acute Myeloid Leukaemia

L – Lymphedema

ID – Immunodeficiency

JMML – Juvenile Myelomonocytic Leukaemia

For patients with multiple *GATA2* mutations, it is not clear whether the mutations exist within individual or separate clones since variant allele frequencies were not reported.

Appendix IV. Supplementary information for Chapter 3.

Manuscript titled ‘Clinically important driver mutations in GATA2 zinc finger 2 display functional diversity and tendency for specification of myeloid malignancy subtypes, immunodeficiency disorders and lymphedema’

Supplemental data

Clinically important driver mutations in GATA2 zinc finger 2 display functional diversity and tendency for specification of myeloid malignancy subtypes, immunodeficiency disorders and lymphedema

Table of Contents

Supplemental materials and methods

Tissue culture and transfection. HEK293, HEK293T and Cos-7 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 1x Penicillin/Streptomycin/Glutamine (Sigma) and 10% FBS (Sigma). All transient transfections were performed as per Lipofectamine 2000 protocol (Invitrogen).

Immunofluorescence staining. HEK293 cells, transiently expressing GATA2, were processed and immunofluorescently stained as described previously.[1]

*Nuclear lysate preparation. m*GATA2 proteins were transiently expressed in HEK293 cells. Nuclear lysates were prepared as described by Andrews and Faller.[2]

Western blotting-electromobility shift assay (WEMSA). WEMSA were performed according to previously published methods.[1]

Electromobility shift assay (EMSA). Purified *h*GATA2 ZF2 proteins were tested for their DNA-binding capacity by EMSA using a GATA site-containing oligonucleotide derived from GATA2 responsive element of the *hGM-CSF* promoter, labelled with either ^{32}P or fluorescein (FAM). Proteins were dialysed into 20 mM Tris (pH 7.9), 50 mM NaCl, 1 mM DTT, and combined at concentrations between 50 nM–6.4 μM with FAM-Prox1 (5 nM) in the above buffer supplemented with 0.2 mg/ml BSA and 5 mM $MgCl_2$. Samples were incubated on ice for 30 min, separated on 8% non-denaturing polyacrylamide gels made in 0.5x TBE (45 mM Tris, 45 mM Boric acid, 2.5 mM EDTA) and visualised using a Typhoon FLA 9000 laser scanner equipped with a 473 nm laser and a LPB filter set (FAM), or by using phosphorimaging (^{32}P) .

GATA2 ZF2 protein production, purification and characterisation. WT and mutant *h*GATA2 ZF2 proteins were expressed and purified as described,[3] with the exception that 20 mM Tris (pH 8.5) was maintained throughout the buffers, and only a single second cation exchange chromatography step, using a 50-600 mM NaCl step gradient was used. ZF2 proteins were dialysed into 20 mM sodium phosphate (pH 7.4), 50 mM NaCl, 1 mM DTT. Far-UV circular dichroism (CD) spectra (195-260 nm) were collected at protein concentrations of 5 μM using a JASCO J-815 CD spectropolarimeter at 25 °C. The resulting spectra were smoothed in Origin (Microcal) using five-point fast Fourier transform filtering. One-dimensional proton (1D-H^1) NMR spectra were collected at 25 °C in the same buffer on a Bruker AvanceIII 800 MHz spectrometer on samples (150-300 μM ZF2 protein) supplemented with 10% (v/v) D_2O, and 17 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid as an internal reference.

Isothermal titration calorimetry (ITC). Wild-type and mutant *h*GATA2 ZF2 proteins (250–270 μM) and *h*GM-CSF oligonucleotide (20 μM) were dialysed into 20 mM Tris (pH 7.8), 150 mM NaCl, 0.5 mM Tris-(2-Carboxyethyl)phosphine (TCEP), and proteins were titrated into the DNA at 25 °C using a MicroCal iTC200. Initial titrations revealed a primary binding event at a 1:1 stoichiometry, followed by an additional weak non-specific binding event, that could not be fitted by a two-site model. Subsequent experiments were designed with additional injections, allowing the last 5-6 points to be fitted with a linear fit to account for non-specific binding, and this line was subtracted from the data. The adjusted data was then fitted to a standard one-site model.

Luciferase reporter assay. Luciferase reporter assays were performed using Dual-Luciferase Reporter Assay System (Promega) using a GloMax®-Multi Detection System (Promega). Experimental methodology is detailed in our previous publication.[1]

Co-immunoprecipitation (Co-IP). Co-IP was performed using Protein G Agarose (Roche) or Pierce™ Crosslink IP Kit (Thermo Scientific). Experiments were carried as per manufacturer's protocol. Briefly, FLAG-tagged GATA2 (WT or mutants) and MYC- or FLAG-tagged PU.1 (NP_003111.2) were co-expressed in HEK293. The cells were then lysed in standard RIPA buffer. Anti-FLAG antibody (M2, Sigma) was used to precipitate FLAG-tagged GATA2, while cross-linked anti-PU.1 antibody (D19; sc-5949, Santa Cruz biotechnology) was used to pull-down PU.1. Western analyses were performed with anti-GATA2 (H116; sc-9008, Santa Cruz biotechnology) or anti-PU.1 antibody.

Retroviral production. Infectious retrovirus was made by co-transfecting HEK293T cells with pMSCV-*m*Gata2-IRES-GFP (WT or mutants) and the packaging plasmid pEQ-Eco (mass ratio 1:1) as described previously.[4] Supernatants were harvested 24 h later and filtered through a 0.45 μm syringe filter (Thermo Fisher Scientific). Viruses were concentrated by ultracentrifugation (*Optima™ XL,* Beckmen Coulter) at 12,100 x g, 4°C for 20 h.

Quantitative real-time PCR (qRT-PCR). For qRT-PCR, 1.5×10^4 LSK cells were lysed using RealTime ready Cell Lysis Kit (Roche) and cDNAs were synthesized using Transcriptor Universal cDNA Master (Roche) as per the manufacturer's instructions. PCR was performed using LightCycler® 480 DNA SYBR Green I Master (Roche) on the LightCycler® 480 Real Time PCR Instrument (Supplementary Table S1).

Cell isolation, retrovirus transduction and clonogenicity assay. Murine hematopoietic cells were isolated from femurs, tibias and hip bones of C57BL/6 mice and resuspended in Iscove's Modified Dulbecco's Medium (Invitrogen) supplemented with 15% FBS. Erythrocytes were purged (RBC lysis buffer, Qiagen) before magnetic-bead depletion of lineage-positive (lin^+) cells (Dynabeads, Invitrogen). $\underline{L}in^-$ $\underline{S}ca1^+$ c-$\underline{K}it^+$ (LSK) cells were then selected using FACS sorting, (streptavidin-APC/Cy7, Scal-PE/Cy7 and c-Kit-APC, BioLegend®, 1:100 each; FACSAria™ II sorter, BD Biocsiences). LSK cells were cultured in retronectin (30 μg/ml, Takara) coated 24-well plates at a concentration of 0.5-1 × 10^5 cells/ml in StemSpan™ SFEM (StemCell Technologies) medium supplemented with SCF, TPO, G-CSF (100 ng/ml each, Peprotech) and Penicillin/Streptomycin/Glutamine (Sigma). Cells were expanded for two days before they were retrovirally spinfected twice (MOI=1 in each round) at 650 ×g, 32 °C for 60 min. The cells were cultured for 2 days before FACS-sorting. GFP^+ LSK cells were then seeded on a methylcellulose medium (MethoCult®, M3434, StemCell Technologies) containing the indicated cytokines. After 7 days of culture, colonies were enumerated and typed under a light microscope. For transplantation assays, 1 × 10^5 GFP^+ LSK cells (transduced with EV, WT, gT354M and gC373R) and 2 × 10^5 helper marrow cells were injected into the tail vein of lethally irradiated (1000 Rad) congenic mice.

Structural modelling of mutant hGATA2 proteins. We have used structural modelling based on the high conservation between ZF2 of human GATA2 and murine GATA3 to gain an insight into why different clinically important mutations alter the ability of GATA2 to bind to DNA.[5] We previously reported disruption to critical interactions for gT354M, sL359V and gR361L.[1,6] Here, we show changes to interactions for R362Q. Using the recently reported structure for human GATA3[7] (PDB ID: 4HCA.1), we can now also model interactions for gR396Q and gR398W (Supplementary Figure S4).

Table S1. Primers for GATA2 mutagenesis, PCR amplification, cloning, EMSA and WEMSA. Nucleotide changes leading to missense mutations (bold underlined). Underlined are the *Eco*RI, *Xho*I, *Nde*I, *Bgl*II and *Xba*I restriction endonuclease sites used for cloning. GATA consensus binding sites (red) and mutated GATA consensus binding sites (red bold underlined). N/A: not applicable.

Primer name	Purpose	Primer sequence (5' → 3')	Ta
*h*R361L-F	Mutagenesis	CACCACCTTATGGC**T**CCGAAACGCCAAC G	55°C
*h*R361L-R	Mutagenesis	CGTTGGCGTTTCGG**A**GCCATAAGGTGGT G	
*h*R362Q-F	Mutagenesis	CACCTTATGGCGCC**A**AAACGCCAACGGG G	
*h*R362Q-R	Mutagenesis	CCCCGTTGGCGTTT**T**GGCGCCATAAGGTG	
*h*C373R-F	Mutagenesis	CTGTCTGCAACGCC**C**GTGGCCTCTACTAC	
*h*C373R-R	Mutagenesis	GTAGTAGAGGCCAC**G**GGCGTTGCAGACA G	
*h*R396Q-F	Mutagenesis	GGGATCCAGACTC**A**GAACCGGAAGATG	
*h*R396Q-R	Mutagenesis	CATCTTCCGGTTC**T**GAGTCTGGATCCC	
*h*R398W-F	Mutagenesis	TCCAGACTCGGAAC**T**GGAAGATGTCCAA C	
*h*R398W-R	Mutagenesis	GTTGGACATCTTCC**A**GTTCCGAGTCTGGA	
*h*GATA2-ZF2-F	PCR & cloning	GTACATATGCAGAACCGACCACTC	58°C
*h*GATA2-ZF2-R	PCR & cloning	GTAAGATCTTCACTATTATTTCTTGCTCTT CTTG	
*m*T354M-F	Mutagenesis	GCACCTGTTGTGCAAATTGTCAGA**T**GAC AACCACCACCTT	55°C
*m*T354M-R	Mutagenesis	AAGGTGGTGGTTGTC**A**TCTGACAATTTGC ACAACAGGTGC	
*m*T355del-F	Mutagenesis	ATTGTCAGACG|ACCACCACCTTATGGCG CCGGAACGCCAACGGGG	
*m*T355del-R	Mutagenesis	CCCCGTTGGCGTTCCGGCGCCATAAGGTG GTGGT|CGTCTGACAAT	
*m*L359V-F	Mutagenesis	ATTGTCAGACGACAACCACCACC**G**TATG GCGCCGGAACGCCAACGGGGACCCTGTG T	
*m*L359V-R	Mutagenesis	ACACAGGGTCCCCGTTGGCGTTCCGGCG CCATA**C**GGTGGTGGTTGTCGTCTGACAAT	
*m*R361L-F	Mutagenesis	CACCACCTTATGGC**T**CCGGAACGCCAAC G	
*m*R361L-R	Mutagenesis	CGTTGGCGTTCCGG**A**GCCATAAGGTGGT G	
*m*R362Q-F	Mutagenesis	ATTGTCAGACGACAACCACCACCTTATG GCGCC**A**GAACGCCAACGGGGACCCTGTG T	

*m*R362Q-R	Mutagenesis	ACACAGGGTCCCCGTTGGCGTTC**T**GGCG CCATAAGGTGGTGGTTGTCGTCTGACAAT	
*m*C373R-F	Mutagenesis	CTGTGTGCAACGCC**C**GTGGCCTCTACTAC	
*m*C373R-R	Mutagenesis	GTAGTAGAGGCCAC**G**GGCGTTGCACACA C	
*m*R398W-F	Mutagenesis	CCAGACCCGGAAT**T**GGAAGATGTCCAG	
*m*R398W-R	Mutagenesis	CTGGACATCTTCC**A**ATTCCGGGTCTGG	
FLAG-*h*GATA2-F	PCR & cloning	GTTCTAGAGCCACCATGGACTACAAGGA TGACGATGACAAGATGGAGGTGGCGCCG GA	65-68°C
FLAG-*h*GATA2-R	PCR & cloning	GTTCTAGACTAGCCCATGGCGGTCACCAT GC	
*m*Gata2(Retro)-F	PCR & cloning	GTAGAATTCGCCACCATGGAGGTGGCGC CTGA	60°C
*m*Gata2(Retro)-R	PCR & cloning	GTACTCGAGCTAGCCCATGGCAGTCACC ATGC	
*m*Gata2-F	qRT-PCR	GCACCTGTTGTGCAAATTGT	
*m*Gata2-R	qRT-PCR	AGCCCCTTTCTTGCTCTTCT	
*m*GAPDH-F	qRT-PCR	CCAATGTGTCCGTCGTGGATC	
*m*GAPDH-R	qRT-PCR	GTTGAAGTCGCAGGAGACAAC	
*h*GMCSF-F	WEMSA & EMSA	TCTCTCGTGATAAAGATCCTGGA	N/A
*h*GMCSF-R	WEMSA & EMSA	TCCAGGATCCTTATCACGAGAGA	
*h*GMCSF-mut-F	WEMSA & EMSA	TCTCTCGT**TC**TAAAGATCCTGGA	
*h*GMCSF-mut-R	WEMSA & EMSA	TCCAGGATCCTTA**GA**ACGAGAGA	
*h*TCRD-F	WEMSA	CACTTGATAACAGAAAGTGATAACTCT	
*h*TCRD-R	WEMSA	AGAGTTATCACTTTCTGTTATCAAGTG	
*h*TCRD-mut-F	WEMSA	CACTT**TC**TAACAGAAAGT**TC**TAACTCT	
*h*TCRD-mut-R	WEMSA	AGAGTTA**GA**ACTTTCTGTTA**GA**AAGTG	
MYC-PU.1-F	PCR & cloning	GTATCTAGAGCCACCATGGAACAAAAAC TTATTTCTGAAGAAGATCTGATGTTACAG G	62°C
MYC-PU.1-R	PCR & cloning	GTATTCTAGATCAGTGGGGCGGGTG	
PU.1-FLAG-F	PCR & cloning	GATTCTAGAGCCGCCATGTTACAGGCGT	
PU.1-FLAG-R	PCR & cloning	GATTCTAGATCACTTGTCGTCATCGTCTT TGTAGTCGTGGGGCGGG	

Table S2. Summary of genetic and clinical details of individuals and families with common germline GATA2 mutations.

Patient ID	Number of patients	GATA2 Mutation Protein/cDNA changes	Overall Phenotype	Lymphedema (age - years)	Hematological abnormalities (age - years)	Co-occurrent Chromosomal Changes	Co-occurrent Acquired Gene Mutations	Immunodeficiency/ Infections (age - years)	Publications
Patient 1 (FHCRC-97)	1 Female (daughter)	p.T354M c.1061C>T	M/A, ID	None	MDS with 5q- clone, aplastic anemia (10)	5q- clone		Recurrent febrile illness, parainfluenza, mycoplasma	6
Patient 2 (FHCRC-84)	1 Female (daughter)	p.T354M c.1061C>T	M/A	None	MDS (14) with 5-6% monosomy 7, thrombocytopenia (14)	-7 (5-6% Blasts)			6
Patient 3 (FHCRC-95)	1 Female (mother)	p.T354M c.1061C>T	M/A, ID	None	MDS (44) with trisomy 8, leukopenia, anemia, thrombocytopenia (15)	+8		Recurrent infections (pneumonias) as a child	6
Kindred 5.II.1	1 Female (mother)	p.T354M c.1061C>T	M/A, ID	None	Cytopenia - T/B/NK/mono (19), MDS (trisomy 8) (34), Pancytopenia (37), increased atypical LGL, progression to AML (41 and 44†)	+8 -6, +r		HPV infection (adolescence) leading to cervical cancer (19), chronic warts, *M. tuberculosis* (32), MAC (32), *M. abscessus* (34)	8-11
Kindred 5.III.1	1 Male (son)	p.T354M c.1061C>T	M/A, ID	None	AML in blast crisis (17 and 19†)			Recurrent HPV infections (child)	9,11

Kindred 17.I.1 (Patient 4)	1 Male	p.T354M c.1061C>T	M, ID	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia, MDS (32)	+8	Disseminated *M.kansasii*(28), Severe HPV (32), Severe *Trychophyton rubrum*(32)	9,10,12
17.II.2	1 Male	p.T354M c.1061C>T	None	None	Unaffected (5)			10
Kindred 19.II.1	1 Male	p.T354M c.1061C>T	ID	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia (20 and 22†)		Chronic EBV infection (20), *M.chelonae*(21)	8-10
Subject 2 (2.I.1)	1	p.T354M c.1061C>T	ID	None	DCML deficiency		HPV, *M.Kansasii*, bacilli and influenza H1N1 infection (27)	13-15
9.III.1	1	p.T354M c.1061C>T	M, ID	None	MDS (31)		HPV, mycobaterial infection (31)	15,16
9.III.2	1	p.T354M c.1061C>T	ID	None	DCML deficiency		HPV (29)	15,16
9.III.3	1	p.T354M c.1061C>T	None	None	Unaffected (22)			15,16
9.III.4	1	p.T354M c.1061C>T	None	None	Unaffected (17)			15,16
9.III.5	1	p.T354M c.1061C>T	M	None	MDS (17)	+8		15,16
I-2	1 Female (great grandmother)	p.T354M c.1061C>T	M/A	None	Leuk (†?)[§]			17
II-2	1 Female (paternal grandmother)	p.T354M c.1061C>T	M/A	None	AML (†?)[§]			17

II-3	1 Male (paternal great uncle)	p.T354M c.1061C>T	M/A	None	Leuk (†?)[§]				17
IV-6	1 Male (son - proband)	p.T354M c.1061C>T	M/A, ID	None	MDS (REAB-2) (23), cytopenias (23)	-7 i 17	***ASXL1*** (p.G646Wfs*12)		17
IV-1	1 Male (cousin)	p.T354M c.1061C>T	M/A, ID	None	MDS (REAB-1) (18), absolute monocytopenia (18)	-7	***ASXL1*** (p.G646Wfs*12)	Recurrent minor infections (<18), warts	17
III-1	1 Female (obligate carrier)	p.T354M c.1061C>T	None	None	Unaffected (60)			None	17
III-5	1 Male (father)	p.T354M c.1061C>T	None	None	Unaffected (52)			None	17
III-7	1 Male (uncle)	p.T354M c.1061C>T	None	None	Unaffected (51)			None	17
IV-10	1 Female (cousin)	p.T354M c.1061C>T	ID	None	Monocytopenia, neutropenia (31), reduced NK cells (personal comm.)			Recurrent minor infections (31), but no childhood infections	17
Family 26	1 Female	p.T354M c.1061C>T	M/A	None	MDS (REAB) (26)				18
Family 26	1 Female (Sister)	p.T354M c.1061C>T	M/A	None	MDS (28)				18
Family 26	1 Female (Mother)	p.T354M c.1061C>T	M/A	None	MDS (24†)				18
Family 26	1 Female (Auntie)	p.T354M c.1061C>T	M/A	None	AML (21†)				18
1	1 Male	p.T354M c.1061C>T	M	None	MDS (33)	+8			19

9	1 Female	p.T354M c.1061C>T	M/A	None	MDS/AML (42†)	dic 6, +8		19
Kindred 35.III.3	1 Male	p.T354M c.1061C>T	M	None	MDS (RCMD) (32)	+8	Granulomatous lymphadenitis (24) and polymicrobial necrotizing fasciitis (28)	10
Pedigree 1 (Adelaide)	7 Males, 8 Females	p.T354M c.1061C>T	M/A	None	6x MDS/AML, 3x AML, 2x AML-M2, 2x Leuk 2x unaffecteds			1
Pedigree 2 (Seattle)	5 Males, 1 Female	p.T354M c.1061C>T	M/A	None	2x MDS, 1x AML, 1x atypical CML, 1x Acute Leuk 1x unaffected			1
Pedigree 3 (Seattle)	6 Males, 6 Females	p.T354M c.1061C>T	M/A	None	2x MDS, 2x MDS/AML, 3x AML-M2, 1x AML M7, 2x Leuk, 1x Macrocytic anemia[§], 1x unaffected			1
Kindred 1.I.2	1 Female (mother)	p.R398W c.1192C>T	M/A like, ID	None	CMML (54), RAEB (54†), Monocytopenia & lymphocytopenia		disseminated *Mycobacterium sp*. (54†)	9,11

Kindred 1.II.1	1 Female (daughter)	p.R398W c.1192C>T	M/A like, ID	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia, CMML (42) blast crisis (43 and 46†)		HPV (20s), HSV (41), EBV+tumour (41), MAC (41,42), *P.jiroveci* (41)	8-11
Kindred 1.II.5 (Patient 2)	1 Female (daughter)	p.R398W c.1192C>T	M,ID	None	MDS (RCMD), increased atypical LGL (36)		HPV (adolescence, 17^), 8 spontaneous abortions, MAC (37), Parvovirus B19 (46), *M.fortuitum* (47), VZV, invasive *Aspergilus* pneumonia (47), *Fusarium fungemia* (47)	9-11,20
Kindred 2.I.1	1 Male (father)	p.R398W c.1192C>T	M/A-like, ID	None	Pancytopenia (42), "leukaemic process" (42)		*M.scrofulaceum* (late 30s, 39^), cryptococcal meningitis, recurrent *staphylococcal* infections	9,11
Kindred 2.II.3	1 Male (son)	p.R398W c.1192C>T	M/A-like, ID	None	Cytopenia - T/B/NK/mono (36), MDS (RAEB-2), increased atypical LGL (38), progression to AML (39†)	+8, +1, der(1;7) (q10;p10)	HPV(20s), Disseminated *H.capsulatum* (34), MAC (35), *Candida glabrata* (39), *Serratia* pneumonia (39), *S. epidermidis* bacteremia	8-11

Kindred 3 3.I.1	1 Female	p.R398W c.1192C>T	M/A, ID	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia (48) CMML (49), increased atypical LGL (59†)		*M.fortuitum* (47,51), MAC (53) invasive aspergillosis, *Scedosporium*, *Scopulariopsis*, *Graphium*, *Pneumocystis* on BAL (59), HPV (59), Fatal *C.difficile* (59)	9-11
Kindred 9.III.1	1 Male	p.R398W c.1192C>T	M, ID	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia (8), MDS (RCMD) (22)		HPV (8), *M.fortuitum*(22)	8-11
Kindred 21.II.1 (Patient 3)	1 Male	p.R398W c.1192C>T	M, ID	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia, MDS (RCMD) (33) BMT (33)	-Y	Disseminated MAC (32), HPV (33), MCV	9,10,20
3.II.6	1	p.R398W c.1192C>T	M, ID	None	DCML deficiency and MDS (34†)		HPV and lung problem (34†)	13-15
3.III.1 (Subject 3)	1 Female	p.R398W c.1192C>T	ID	None	DCML deficiency BMT (23)		HPV and lung problem (22)	13-15
3.III.3	1	p.R398W c.1192C>T	None	None	Unaffected (26)			13-15

5.I.1	1	p.R398W c.1192C>T	M, ID	None	DCML deficiency and MDS (40†)			HPV and mycobacterial infection (40†)	13-15
3	1 Female (sister)	p.R398W c.1192C>T	M/A-like	None	CMML (49†)		***ASXL1*** (p.G646fs – c.1934insG)		19
4	1 Female (sister)	p.R398W c.1192C>T	M/A-like	None	CMML (46†)		***ASXL1*** (p.R693X – c.2077C>T)		19
8	1 Female	p.R398W c.1192C>T	M/A-like	None	CMML (59†)		***ASXL1*** (p.G646fs – c.1934insG)		19
35	1 Male	p.R398W c.1192C>T	M	None	MDS (22)				19
47	1 Male	p.R398W c.1192C>T	M	None	MDS (33)	-Y			19
Kindred 18.I.2 (Patient 4)	1 Female	p.R396Q c.1187 G>A	ID, M	None	Monocytopenia & B-cell, NK-cell and T-cell lymphocytopenia, Atypical megakaryocytes, MDS (RCMD) (16), BMT (23)	+8		Disseminated *M.abscessus* (22) and nocardiosis (22).	9,10,20
Japanese patient	1 Female	p.R396Q c.1187 G>A	L, M/A, ID	Lymphedema (Emberger) (13)	Monocytopenia (4) & B-cell (19), NK-cell (19) and DC deficiency (19), neutropenia (4), MDS (trilineage) (19)			Severe VZV infection (2), recurrent lung infections (>2), *Salmonella enterocolitis* (8), warts (19)	21

Family 23	1 Male	p.R396Q c.1187 G>A	M/A	None	AML-M2 (14)				18
Family 23	1 Male (brother)	p.R396Q c.1187 G>A	M/A	None	MDS (16)				18
Family 23	1 Female (mother)	p.R396Q c.1187 G>A	M/A	None	AML-M2 (?†)[§]				18
16	1 Female	p.R396Q c.1187 G>A	M	None	MDS (23)	+8			19
22	1 Male	p.R396Q c.1187 G>A	ID	None	Low B cell (55)				19
26	1 Male (brother)	p.R396Q c.1187 G>A	M	None	MDS (32)				19
27	1 Male (brother)	p.R396Q c.1187 G>A	M	None	MDS (29)				19
46	1 Female	p.R396Q c.1187 G>A	M	None	MDS (12)				19
14.I.1	1 Female	p.R396Q c.1187 G>A	ID	None	Monocytopenia & B-cell, NK-cell deficiency (7)			EBV (7), HSV (7), MRSA (7)	10,11
31.II.1	1 Male	p.R396Q c.1187 G>A	ID, M	None	MDS (RCMD) (31)			HPV (29), Granulomatous lymphadenitis (27)	10
31.II.2	1 Male	p.R396Q c.1187 G>A	M	None	MDS (RCMD) (29)				10
40.I.1	1 Male	p.R396Q c.1187 G>A		None	Megakaryocyte atypia (55)				10
40.II.1	1 Male	p.R396Q c.1187 G>A	ID, M	None	MDS (RCMD) (18)	+8		HPV (18), VZV	10
40.II.2	1 Male	p.R396Q c.1187 G>A	ID, M	None	MDS (RCMD) (15)	+8		HPV (16)	10

40.II.3	1 Male	p.R396Q c.1187 G>A	ID, M	None	MDS (RCMD) (15)	+8	HPV (14), VZV	10
P1.II.4 (#6227)	1 Female (Mother)	p.R396Q c.1187 G>A	M/A, ID	None	Chronic neutropenia and monocytopenia (23), MDS/AML-M2 (35)		Chronic infections - cutaneous and oral mucosa (23), CMV pneumonia (37†)	22
P1.III.2 (#6165)	1 Male (Son)	p.R396Q c.1187 G>A	M/A, ID	None	Intermittent neutropenia, monocytopenia & B-cell, NK-cell deficiency, RCMD (16)		Chronic weak EBV, severe pneumonia (16)	22
P1.III.3 (#6225)	1 Male (Son)	p.R396Q c.1187 G>A	M/A, ID	None	Mild neutropenia, MDS/AML-M2 (14), BMT (14)	+11, der 7	Recurrent infections (12)	22
P1.III.5 (#6224)	1 Male (Son)	p.R396Q c.1187 G>A	M/A, ID	None	MDS with monosomy 7 (6), neutropenia (6), BMT (6)	-7	Chronic weak EBV	22

Abbreviations: † = Deceased, ^ = assumed age, § = excluded from statistical analysis, +1 = Trisomy 1, +11 = Trisomy 11, +8 = Trisomy 8, +r = Ring chromosome, -7 = Monosomy 7, -Y = loss of Y chromosome, 5q- = Chromosome 5q deletion syndrome, 7q- = Chromosome 7q deletion syndrome, AML = Acute myeloid leukemia, AML-M2 = AML with maturation (M2), AML-M7 = Acute megakaryocytic leukemia, BMT = Bone marrow transplantation, CMML = Chronic myelomonocytic leukemia, CML = Chronic myeloid leukemia, CMV = Cytomegalovirus, dic = Dicentric, der = Derivative chromosome, HPV = Human papillomavirus, HSV = Herpes simplex virus, i(17) = Isochromosome 17, ID = Immunodeficiency, L = Lymphedema, Leuk = Leukemia, LGL = Large granular lymphocyte, M = MDS, M/A = MDS/AML, MAC = Mycobacterium avium complex, MCV = Molluscum contagiosum virus, MDS = Myelodysplastic Syndrome, MRSA = Methicillin-resistant *Staphylococcus aureus*, RAEB = Refractory anemia with excess blasts, RCMD = Refractory cytopenia with multilineage dysplasia, T/B/NK/mono = T-cells/B-cells/Natural killer cells/Monocytes, VZV = Varicella zoster virus.

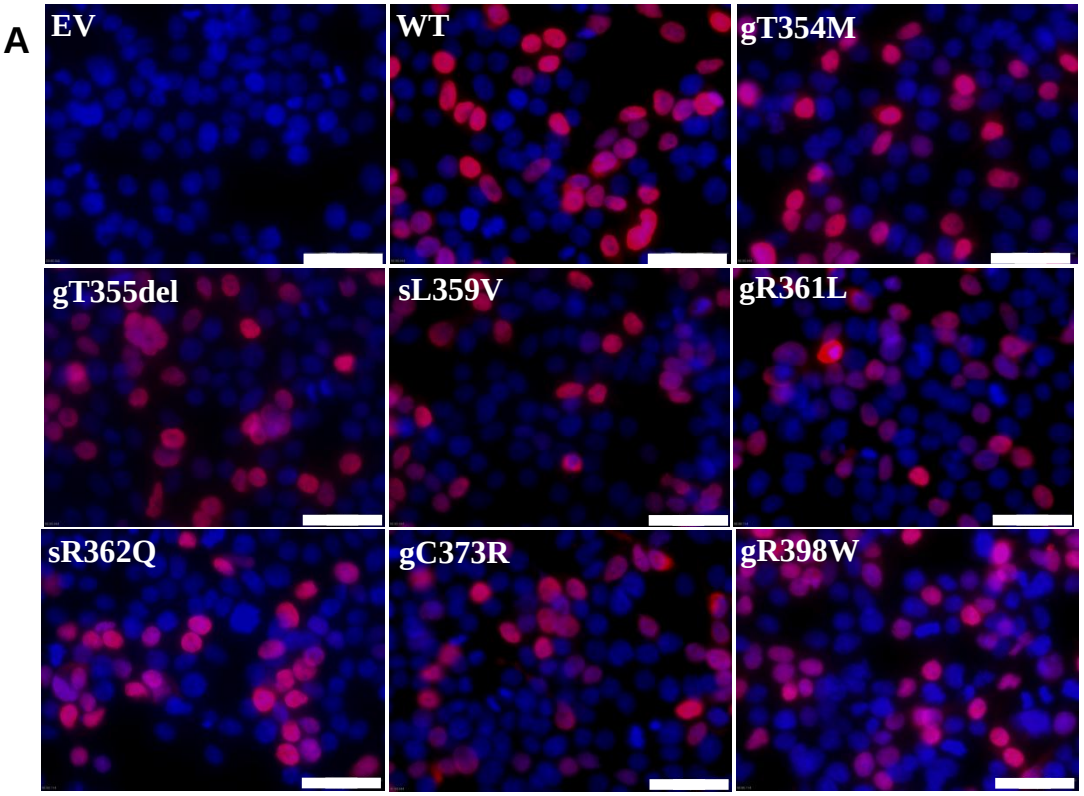
Table S3. Flow cytometric analysis of GFP expression in peripheral blood cells. Peripheral blood (PB) from bone marrow transplanted mice at various time points post-transplantation was obtained for full blood analysis and GFP expression analysis by FACS. All transplanted mice displayed no significant abnormality in full blood counts and blood cell morphology (data not shown). The efficiency of hematopoietic engraftment in recipient mice was determined by analysing leukocytes in the peripheral blood samples for GFP expression. Overall, LSK cells transduced with empty vector (EV) reconstituted better than other constructs as indicated by the good percentage of GFP^+ cells achieved in the PB of some transplanted mice. In contrast, low percentage of GFP^+ circulating leukocytes were generally detected in mice transplanted with GATA2 WT and mutants LSK cells, suggesting that GATA2 expression may suppress long term engraftment in these transplanted mice. The suppressive effect was most pronounced in GATA2 WT, followed by gT354M and gC373R. This result is concordant with *in vitro* functional assays in which expression of GATA2 WT inhibits cell growth in colony forming assays while gT354M and gC373R exhibit LOF characteristics allowing colony outgrowth. Although several mice carrying GATA2 mutants were obtained, none developed acute blood disorders. The average percentage of GFP^+ leukocytes for all transplanted mice remained stable in subsequent PB analysis (data not shown) implicating that long-term repopulating stem cells successfully engrafted into these mice. **Shown are the results* for first bleed.

Construct	Unique ID	GFP^+ cells in PB	PB analysis post-transplantation*	Transplanted mice with GFP^+ cells in PB (GFP^+/total)
EV (n = 11)	1a1	93.8%	15 weeks	100% (11/11)
	1a2	2.6%	15 weeks	
	2d1	2.6%	15 weeks	
	3a2	1.2%	10 weeks	
	3d1	32.5%	10 weeks	
	4c2	95.7%	6 weeks	
	4d2	17.2%	6 weeks	
	4d3	43%	6 weeks	
	5b4	30.7%	4 weeks	
	5b5	46.7%	4 weeks	
	5b6	42%	4 weeks	
WT (n = 11)	1b1	0	15 weeks	36.3% (5/11)
	1b2	0	15 weeks	
	1b3	0	15 weeks	
	3b3	1	10 weeks	
	3d3	0	10 weeks	
	3d4	0	10 weeks	
	4a1	7.8%	6 weeks	
	5b1	1.5%	4 weeks	
	5b2	2.4%	4 weeks	
	5b3	1.2%	4 weeks	
	5b5	0	4 weeks	

gT354M (n = 9)	1c1	88.5%	15 weeks	44.4% (4/9)
	1c2	0	15 weeks	
	1c3	0	15 weeks	
	2c1	0	15 weeks	
	3b2	1.1%	10 weeks	
	3b5	0	10 weeks	
	3c1	0	10 weeks	
	5d1	8.7%	4 weeks	
	5d4	5.6%	4 weeks	
gC373R (n = 7)	2c3	11.7%	15 weeks	71.4% (5/7)
	3c2	2%	10 weeks	
	3c3	20.5%	10 weeks	
	3d2	1.6%	10 weeks	
	5b3	1.2%	4 weeks	
	5c3	0	4 weeks	
	5c4	0	4 weeks	

Supplemental Figures

Figure S1. **GATA2 mutants appropriately localize to nucleus.** Mouse *Gata2* WT or mutant constructs (A) and the human counterparts (B) were transfected into HEK293 and immunofluorescence-stained using anti-GATA2 antibodies and Alexa 555-labelled secondary antibodies. The cells were stained for GATA2 (pink) and DAPI (blue). Scale bars, 50 μm.



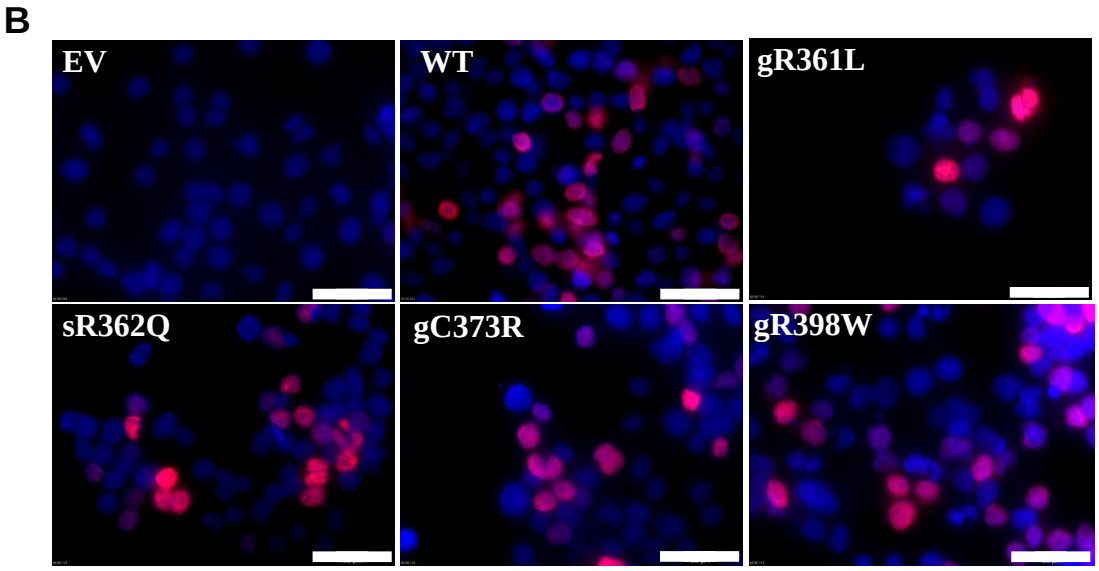


Figure S2. Human GATA2 WT and mutant proteins exhibit differential DNA binding affinity. Full length of human GATA2 WT or mutant proteins was expressed in HEK293 cells. (A) The resultant nuclear extracts were used for Western analysis. (B) 10 μg each of the nuclear lysate was mixed with human *TCRD* enhancer oligonucleotide probe and then used for DNA binding assay. Oligonucleotide probes, C, G and Mut were used as control. C: GATA consensus oligonucleotide; G: *hGM-CSF* enhancer and Mut: GATA binding site mutated *hTCRD* enhancer. Note that a similar DNA binding pattern was found for murine GATA2 WT and mutant proteins (see Figure 1).

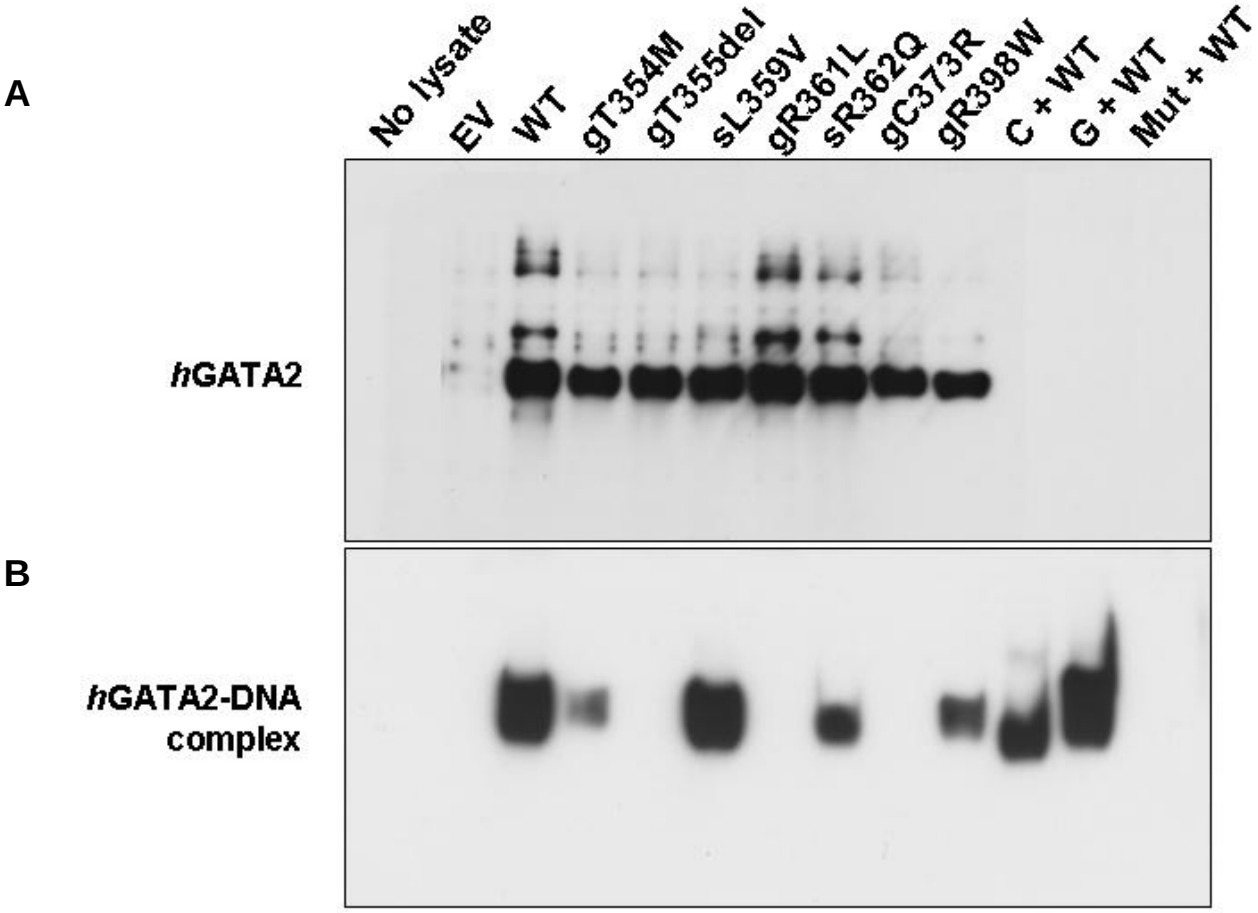


Figure S3. GATA2 mutants differ in their DNA binding affinity when quantified by electromobility shift assay (EMSA) and isothermal titration calorimetry (ITC). *h*GATA2 ZF2 WT and mutant constructs were expressed and purified, and the effect of the mutation on their DNA-binding affinity was assessed. For this study, we also included the gR396Q, a mutant associated with congenital neutropenia,[22] MonoMAC,[9,11] MDS/AML[10,18,19] and Emberger syndrome.[21] (Left) The ZF2 mutants were first compared using EMSA where the ZF2 domains were titrated onto an oligonucleotide derived from the *hGM-CSF* enhancer, labelled with either ^{32}P (A) or fluorescein (B). (Centre) The average dissociation constant (K_d) for the *hGM-CSF* oligonucleotide was then determined using ITC (average ± S.E.M where more than two titrations were conducted; see also Supplemental Figure 5,6). (Right) Dissociation constants (K_d) for WT and mutants (Mut) were then compared as a ratio to demonstrate magnitude of loss of DNA binding affinity. ND: not determined. K_d could not be obtained using ITC for two mutants (data not shown); gR361L because the observable binding was too weak to be practicably measured using this technique, and gC373R because the protein displayed abnormal behaviour upon dilution in control experiments that obscured any possible measurement of DNA binding.

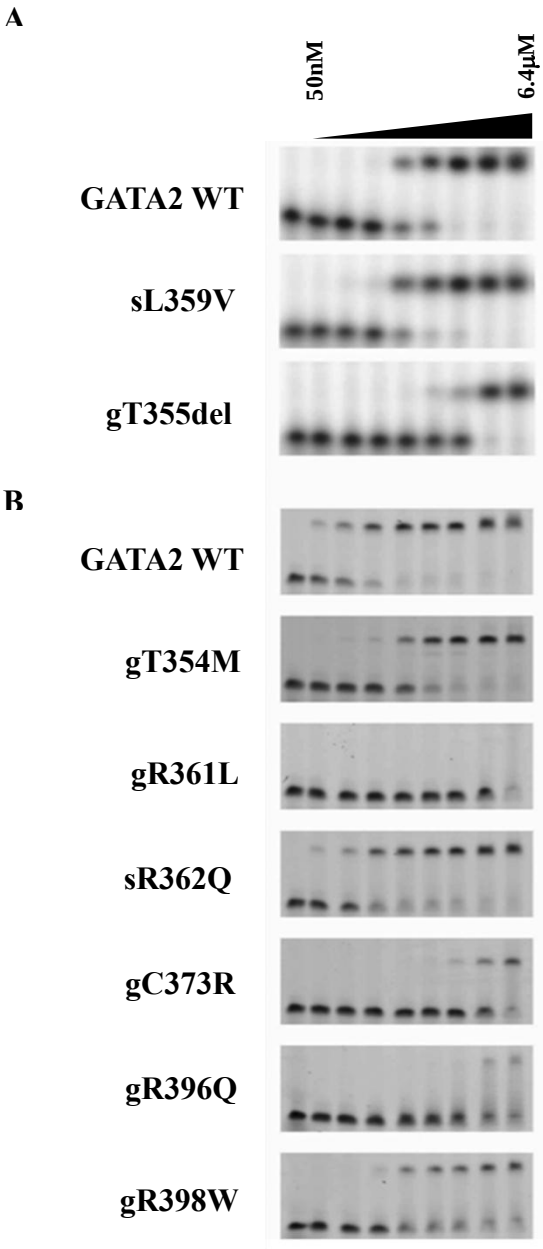
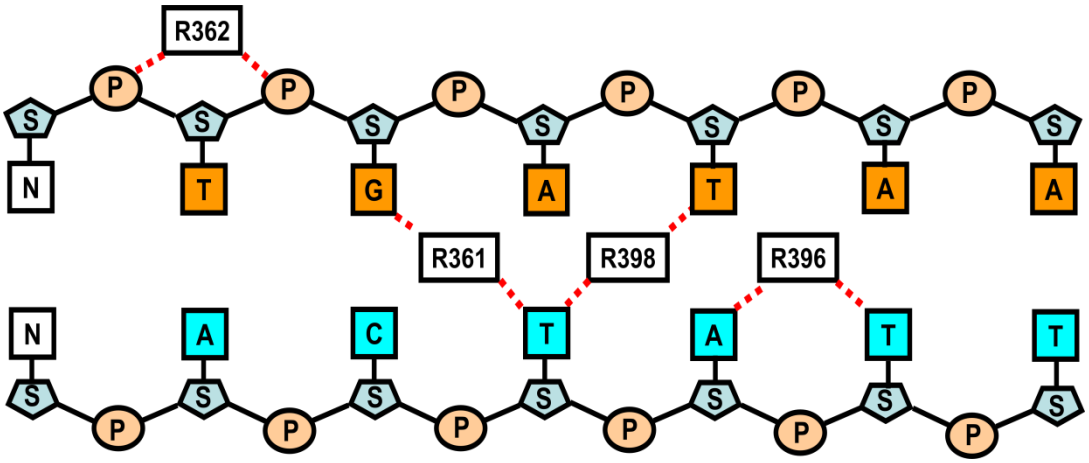


Figure S4. Structural modelling of *h*GATA2. (A) A schematic summary for the conserved arginine residues in the GATA2 ZF2 domain which interact with WGATAR consensus core sequence. The homology modelling (see B) indicates that mutation of positively charged arginine residues (R361 (black), R362 (red), R396 (yellow) and R398 (green)) likely disrupts critical interactions with DNA DNA:protein residue interaction (red dotted line), P: phosphate group and S: sugar. **(B) Protein structure modelling.** Homology modelling was performed to predict DNA binding potential of human GATA2 mutants by using human GATA3 (PDB ID: 4HCA.1) as template. GATA2 protein structure modelling predicts that R361 is an important residue that recognizes the GATA2 consensus WGATAR DNA motif in the major groove by interlocking both sense and antisense bases of the double stranded DNA, while R396 interacts with adjacent thymine and adenine residues of the consensus WGATAR DNA motif. Whereas, R362 and R398 seem to serve as auxiliary residues that reinforce DNA interaction by binding to the negatively charged phosphates of the DNA backbone and/or contributing to ZF2 specificity in the DNA minor groove. The ZF2 domain (palegreen), zinc ion (white sphere), WT arginine residues mutated in each panel (R361 (black), R362 (red), R396 (yellow) and R398 (green)), bases or phosphate group contacted by the mutated arginine (mesh), residues that interact with the mutated arginine using contact changed by the mutation (white), R361L, R362Q, R396Q, R398W (hotpink), major groove (AGATAA sense strand, orange; TTATCT antisense strand, cyan) and critical interactions (black dotted lines).



B

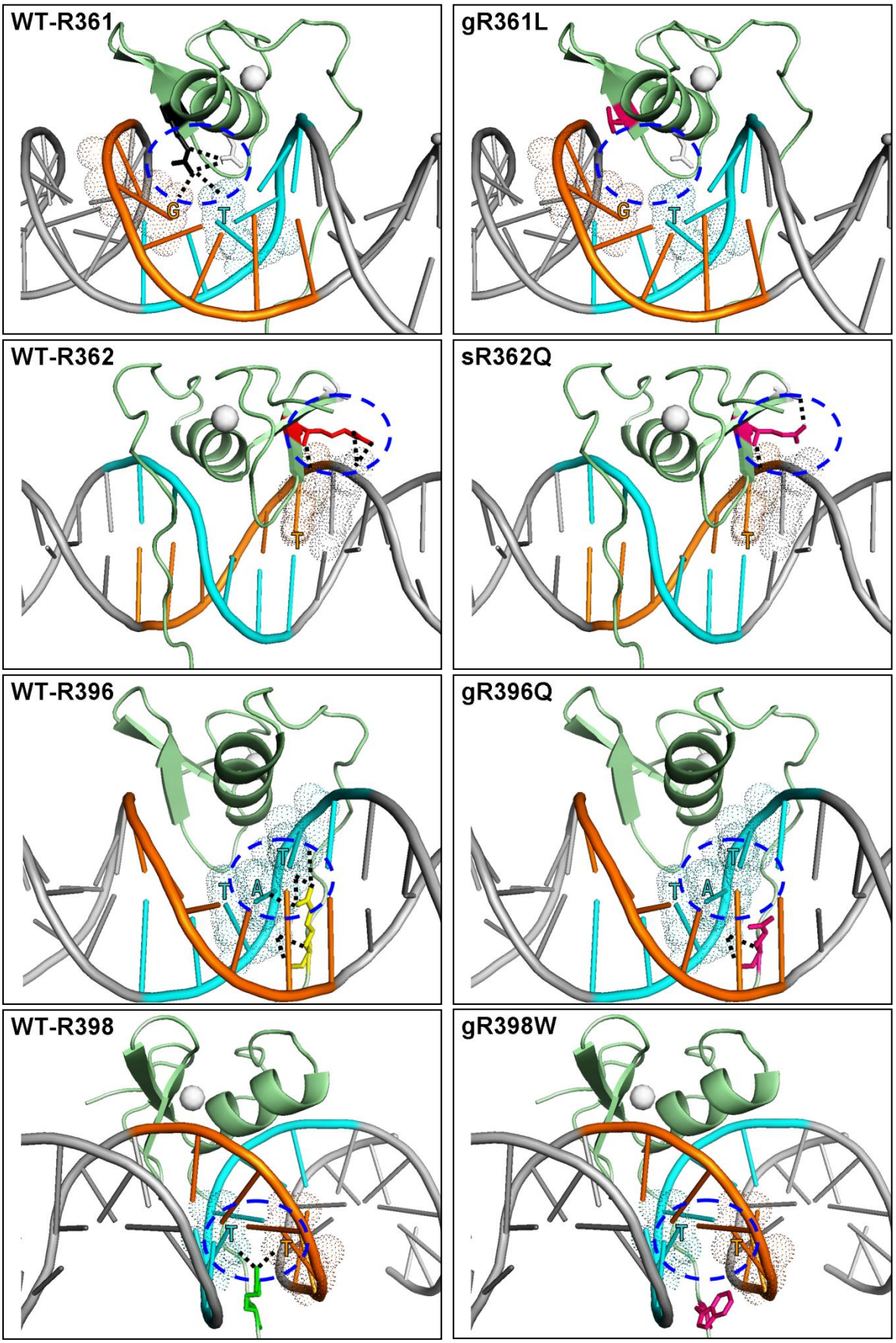

WT-R361
G
T
gR361L
G
T
WT-R362
T
sR362Q
T
WT-R396
T
T
A
gR396Q
T
T
A
WT-R398
T
T
gR398W
T
T

Figure S5. Isothermal titration calorimetry data for WT and mutant *h*GATA2 ZF2 domains. Isothermal titration calorimetry (ITC) was conducted by titrating purified ZF2 domain (250-270 μM) into the *hGM-CSF* GATA2-responsive element oligonucleotide (20 μM) at 25°C. Shown are example titrations for WT and all mutant ZF2 domains for which ITC could be successfully performed.

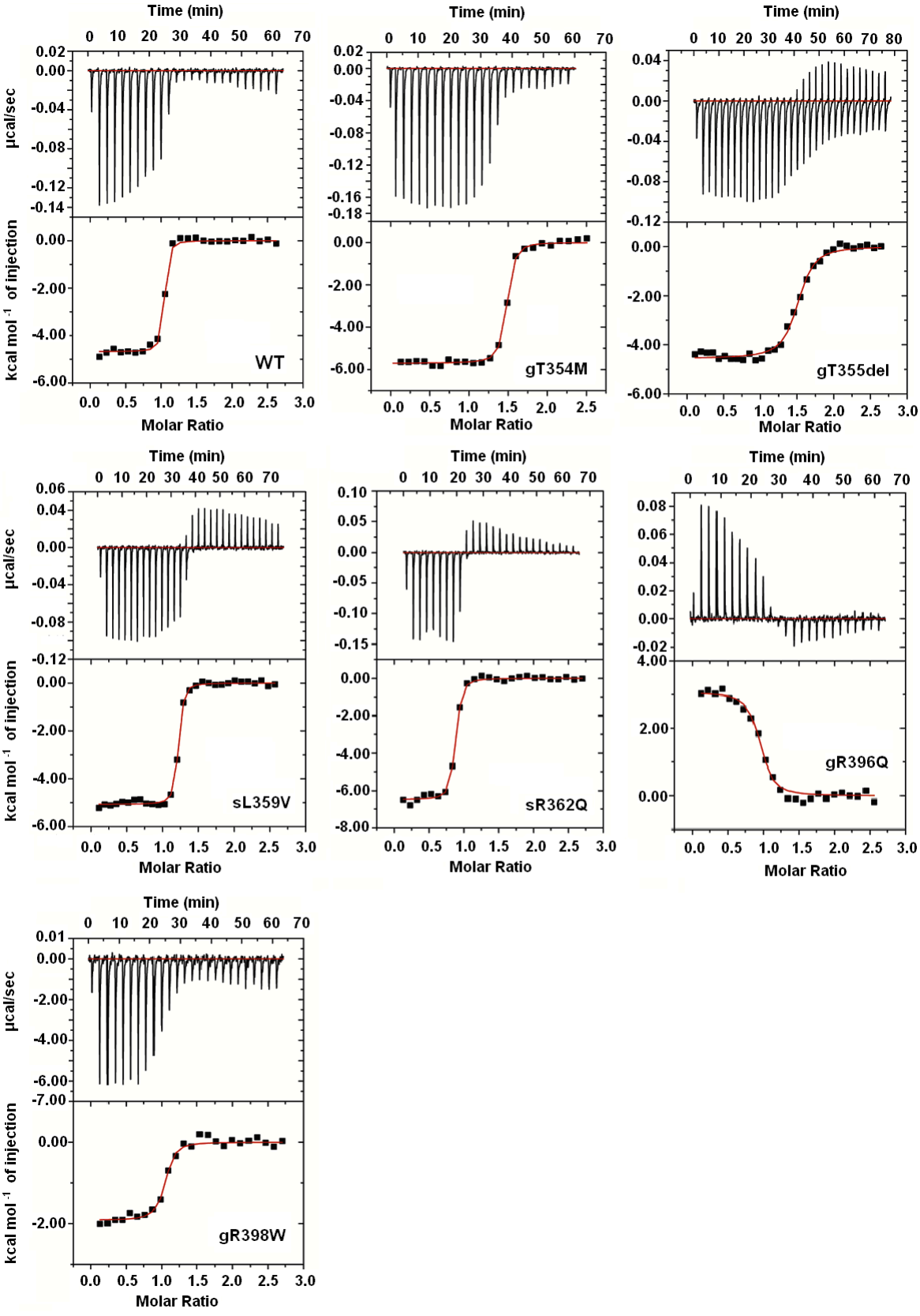
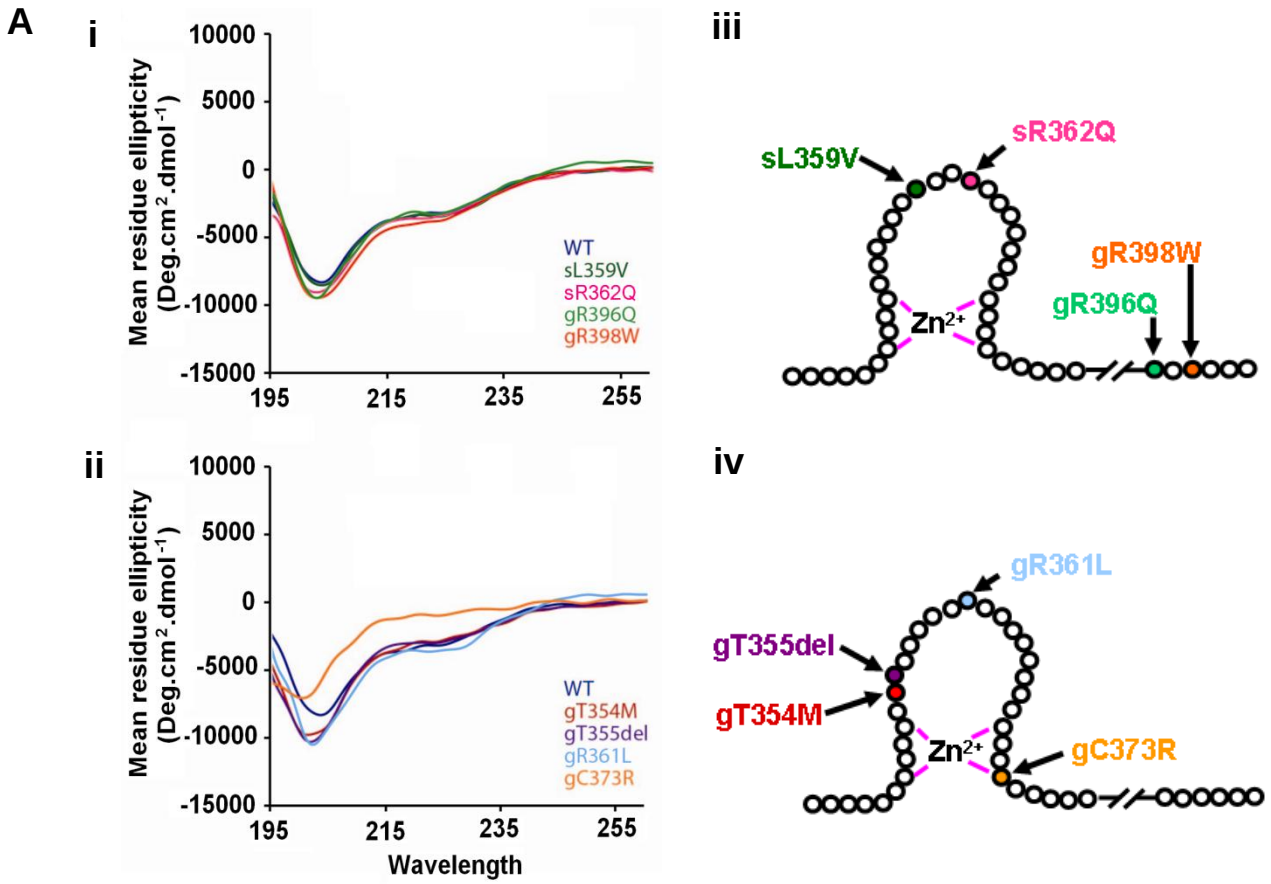


Figure S6. Characterization of the folding of WT and mutant *h*GATA2 ZF2 domains. (A) Circular dichroism (CD) spectra of purified ZF2 domains, grouped by those mutants with a CD spectra similar to WT (Ai), those somewhat different to WT (Aii) and their mutated residues in the ZF2 domain (Aiii,iv). Overall, CD data indicates that all GATA2 mutants except for gC373R have similar but non-identical levels of WT secondary structure. (B) 1D ^{1}H NMR spectra (amide region) of all ZF2 proteins. The sharp well-dispersed peaks seen for WT and many of the mutants indicate a folded protein domain, while the similarity between those spectra suggests that the mutant proteins but gC373R have WT-like levels of folded structure. The tight clustering of peaks around 7-8 ppm for gC373R is characteristic of an unfolded protein. The spectra for gT354M and gT355del show broader peaks and intermediate levels of peak dispersion, which are characteristic of a partially folded protein that may be undergoing low levels of aggregation. Meanwhile, the data for sL359V, gR361L, gR362Q, gR396Q and gR398W suggest that the tertiary structure for these mutants is not significantly affected. Note that peaks >10 ppm correspond to a proton on the tryptophan (W) side chain; all variants contain one W residue except for gR398W, which contains two (see Kwan *et al*. 2011[23]).



B

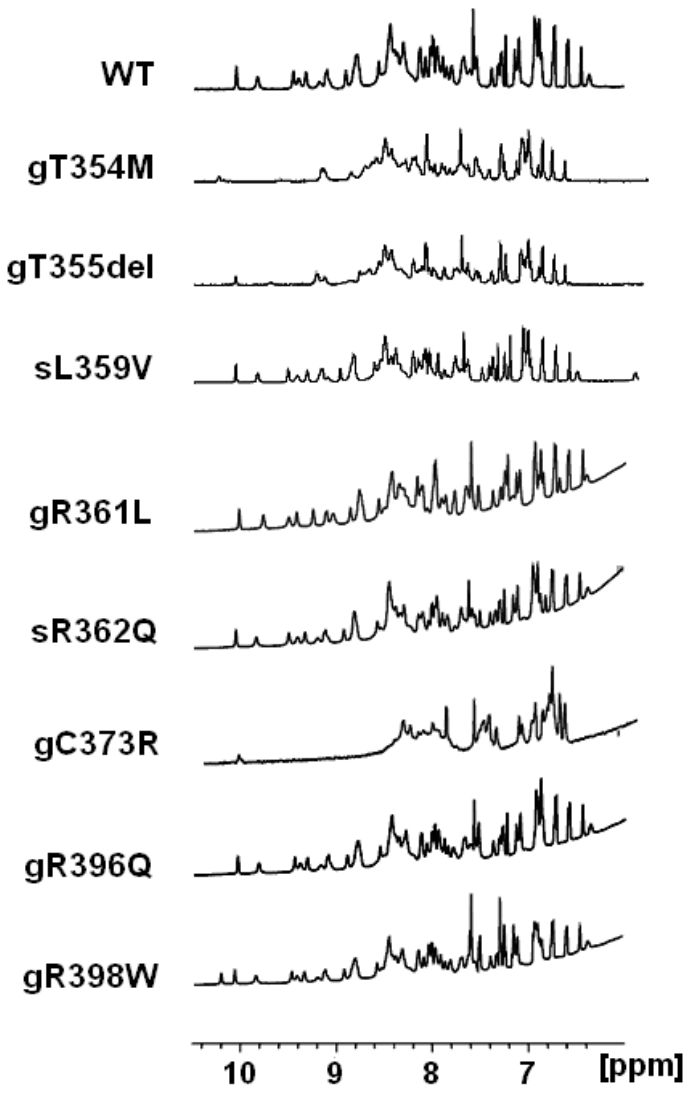
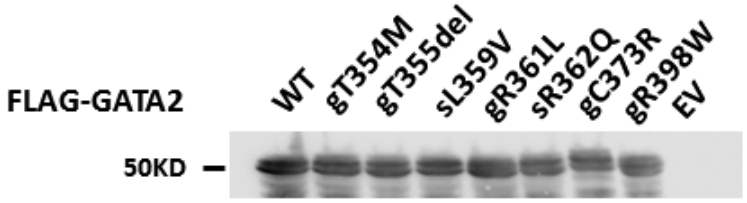
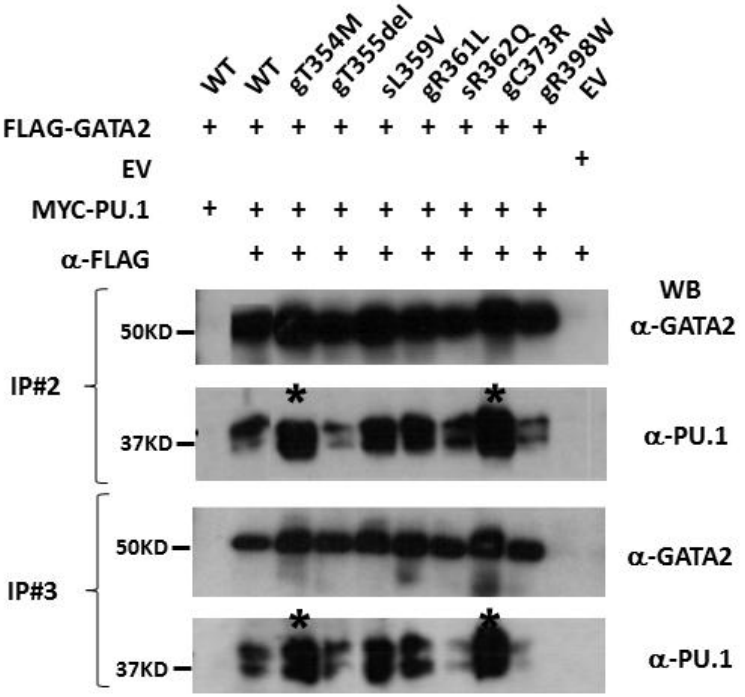

WT
gT354M
gT355del
sL359V
gR361L
sR362Q
gC373R
gR396Q
gR398W
10
9
8
7
[ppm]

Figure S7. Co-immunoprecipitation of GATA2 WT or mutant proteins and PU.1. (A) Expression of GATA2 proteins in HEK293 cells. All FLAG-GATA2 constructs expressed a comparable protein level in the absence of PU.1 (see also Figure 4 for GATA2/PU.1 co-expression). (B) Immunoprecipitation of GATA2 led to co-precipitation of PU.1. (C) GATA2 co-precipitated with PU.1 in reverse IP. gT354M and gC373R (indicated by *) consistently demonstrated enhanced protein-protein interaction in all co-IP experiments. Input lysates not shown.

A



B



C

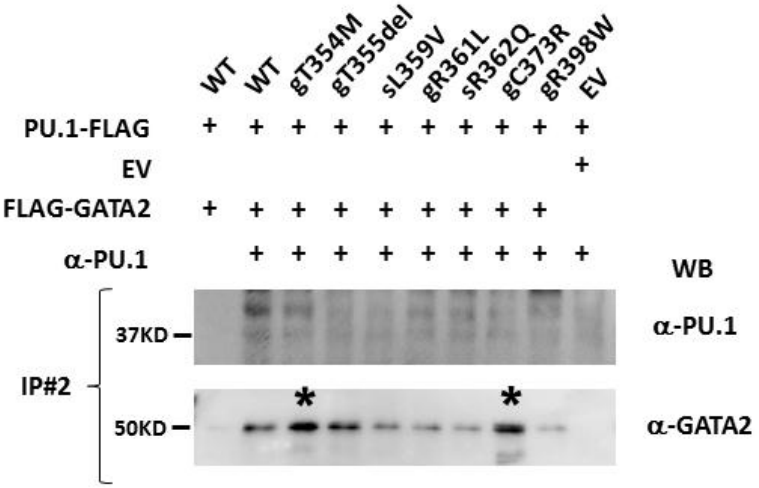


Figure S8. Expression levels of *mGata2* transcript in LSK cells. Quantification of *mGata2* mRNA expression levels in GFP positive LSK cells was assessed by quantitative RT-PCR. Results were normalized to LSK cells transduced with empty vector (EV, set at 1). Data is representative of three independent experiments.

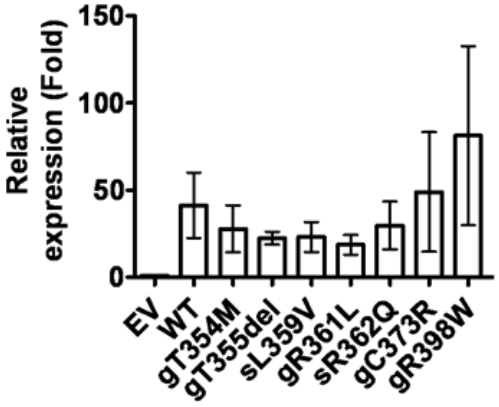
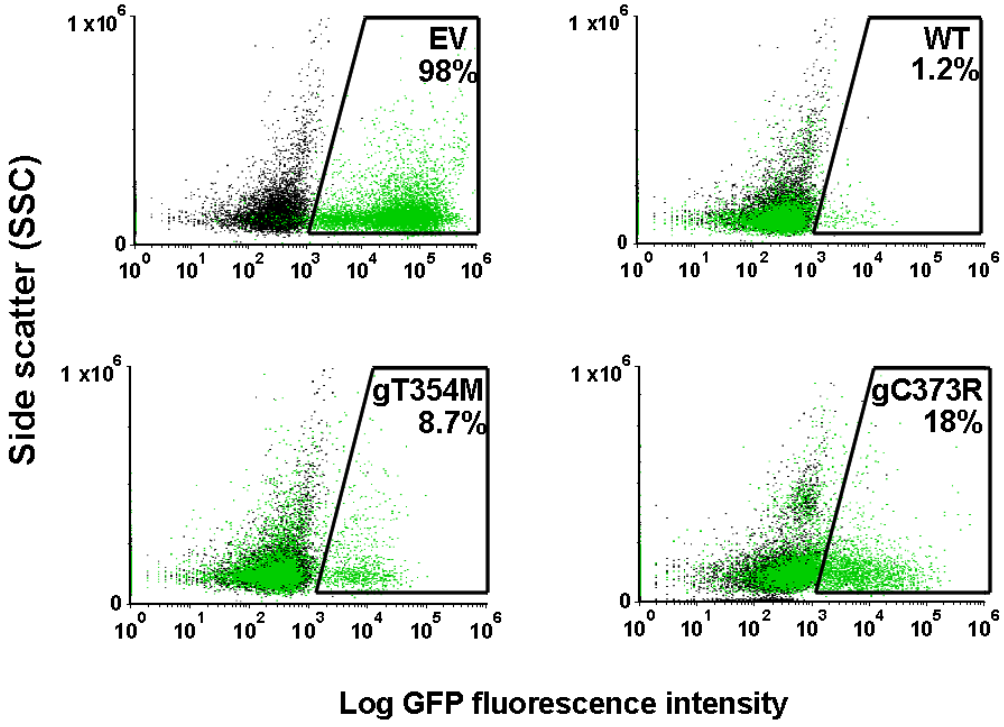


Figure S9. Flow cytometric analysis of GFP expressing cells in the peripheral blood of transplanted mice. Flow cytometric dot plots of peripheral blood from representative mice transplanted with LSK cells transduced with empty vector, *Gata2* wildtype or mutant constructs (green dots) compared with untransplanted host mice (black dots) to determine the level of engraftment of transduced cells. Peripheral blood analysis was performed at 4 weeks (WT and gT354M), 6 weeks (EV) and 10 weeks (gC373R) post transplantation. The percentage of GFP positive cells for each construct is shown in the trapezium gate (only representative results for the first bleed are shown).



References

1. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*. 2011;43(10):1012-1017.
2. Andrews NC, Faller DV. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res*. 1991;19(9):2499.
3. Wilkinson-White L, Gamsjaeger R, Dastmalchi S, et al. Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. *Proc Natl Acad Sci U S A*. 2011;108(35):14443-14448.
4. Brown AL, Peters M, D'Andrea RJ, Gonda TJ. Constitutive mutants of the GM-CSF receptor reveal multiple pathways leading to myeloid cell survival, proliferation, and granulocyte-macrophage differentiation. *Blood*. 2004;103(2):507-516.
5. Bates DL, Chen Y, Kim G, Guo L, Chen L. 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):1292-1306.
6. Kazenwadel J, Secker GA, Liu YJ, 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):1283-1291.
7. Chen Y, Bates DL, Dey R, et al. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep*. 2012;2(5):1197-1206.
8. Hsu AP, Johnson KD, Falcone EL, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood*. 2013;121(19):3830-3837, S3831-3837.
9. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011;118(10):2653-2655.
10. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*. 2014;123(6):809-821.
11. Vinh DC, Patel SY, Uzel G, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood*. 2010;115(8):1519-1529.
12. Mace EM, Hsu AP, Monaco-Shawver L, et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset. *Blood*. 2013;121(14):2669-2677.
13. Bigley V, Haniffa M, Doulatov S, et al. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J Exp Med*. 2011;208(2):227-234.
14. Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*. 2011;118(10):2656-2658.
15. Dickinson RE, Milne P, Jardine L, et al. The evolution of cellular deficiency in GATA2 mutation. *Blood*. 2014;123(6):863-874.

16. Kaur J, Catovsky D, Valdimarsson H, Jensson O, Spiers AS. Familial acute myeloid leukaemia with acquired Pelger-Huet anomaly and aneuploidy of C group. *Br Med J*. 1972;4(5836):327-331.
17. Bodor C, Renneville A, Smith M, 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):890-894.
18. Holme H, Hossain U, Kirwan M, Walne A, Vulliamy T, Dokal I. Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *Br J Haematol*. 2012;158(2):242-248.
19. West RR, Hsu AP, Holland SM, Cuellar-Rodriguez J, Hickstein DD. Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. *Haematologica*. 2014;99(2):276-281.
20. Cuellar-Rodriguez J, Gea-Banacloche J, Freeman AF, et al. Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency. *Blood*. 2011;118(13):3715-3720.
21. Ishida H, Imai K, Honma K, 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):1273-1276.
22. Pasquet M, Bellanne-Chantelot C, Tavitian S, 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):822-829.
23. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J*. 2011;278(5):687-703.

Appendix V. Antibodies Used for Animal Work

Lineage depletion:

Biotin anti-mouse Gr1 (Clone RB6-8C5; Biolegend Cat 108404)

Biotin anti-mouse Ter119 (Clone TER-119; Biolegend Cat 116204)

Biotin anti-mouse B220 (Clone RA3-6B2; Biolegend Cat 103204)

Biotin anti-mouse Cd3 (Clone 145-2C11l; Biolegend Cat 100304)

Biotin anti-mouse Cd5 (Clone 53-7.3; Biolegend Cat 100604)

Biotin anti-mouse Mac1 (Clone M1/70; Biolegend Cat 101204)

Dynabeads Biotin Binder from Invitrogen (Cat no 110.47) for Lineage depletion

Antibodies for LSK sort:

APC/Cy7 Streptavidin (Biolegend Cat 405208)

APC anti-mouse CD117 (c-kit) (Clone 2B8; Biolegend Cat 105812)

PE/Cy7 anti-mouse Ly-6A/E (Sca-1) (Clone D7; Biolegend Cat 108114)

Antibodies for Immunohistochemistry:

Biotin anti-mouse B220 (Clone RA3-6B2; Cat 103204)

Biotin anti-mouse Cd3 (Clone 145-2C11l ; Cat 100304)

Anti-mouse F4/80 Antigen Purified (Clone BM8; eBioscience Cat 14-4801)

Anti-mouse Galectin-3 (MAC2) Purified (Clone eBioM3/M38; eBioscience Cat 14-5301)

Appendix VI. Description of genes whose expression was studied in AML samples in Chapter 4.

Gene name	Description	Expression regulated in MDS or AML	Mutated in MDS or AML	Role in haematopoiesis	GATA2 target	Regulator of GATA2	GATA2 interacting partner	References
ALS2CR8	Amyotrophic Lateral Sclerosis 2 (Juvenile) Chromosome Region, Candidate 8	x						(103)
ANGEL1 (aka KIAA0759)	Angel Homolog 1 (Drosophila)	x						(103)
AR	Androgen receptor				x		x	(195)
ARL6IP5	ADP-Ribosylation Factor-Like 6 Interacting Protein 5	x						(103)
ASXL1	Additional sex combs like 1		x					(196)
B2M	Beta-2-microglobulin							
BAALC	Brain And Acute Leukemia, Cytoplasmic	x						(197)
BAP1	BRCA1 associated protein 1	x	x					(198)
BMI1 (aka PCGF4)	BMI1 proto-oncogene, polycomb ring finger	x						(199)
BSPRY	B-Box And SPRY Domain Containing	x						(103)
BTBD3	BTB (POZ) Domain Containing 3	x						(103)

C1RL	Complement Component 1, R Subcomponent-Like	x						(103)
CD34	CD34 molecule	x						(200)
CEBPA	CCAAT enhancer binding protein alpha		x	x	x	x	x	(201)
CPT1A	Carnitine Palmitoyltransferase 1A (Liver)	x						(103)
CSF2	Colony Stimulating Factor 2 (Granulocyte-Macrophage)	x		x	x			(202, 203)
CSF2RB (βcommon)	Colony stimulating factor 2 receptor beta common subunit	x	x					(204)
CTCF	CCCTC-binding factor		x					(178, 205)
DAPK1	Death-Associated Protein Kinase 1	x						(103)
DNMT3A	DNA methyl transferase 3A		x					(206)
ERG	V-Ets Avian Erythroblastosis Virus E26 Oncogene Homolog	x		x	x	x		(127)
ETFB	Electron-Transfer-Flavoprotein, Beta Polypeptide	x						(103)
EVI2A (EVI2)	Ecotropic Viral Integration Site 2A	x						(207)
EZH2	Enhancer of Zeste Homologue 2		x					(208)
FGFR1	Fibroblast Growth Factor Receptor 1	x						(103)
FLI1	Friend leukemia virus integration 1			x			x	(129)

FLT3	Fms related tyrosine kinase 3		x					(209)
FOSB	FBJ murine osteosarcoma viral oncogene homolog B	x						(210)
GATA1	GATA binding protein 1 (globin transcription factor 1)	x		x	x	x	x	(211, 212)
GATA2	GATA Binding Protein 2	x	x	x	x	x		(19, 58)
GFI1B	Growth Factor Independent 1B Transcription Repressor	x	x			x		(213, 214)
HEATR6 (aka ABC1)	HEAT Repeat Containing 6	x						(103)
HMBS	Hydroxymethylbilane Synthase							
HOXA9	Homeobox A9	x				x		(215)
HPRT1	Hypoxanthine Phosphoribosyltransferase 1							
ID1	Inhibitor Of DNA Binding 1, Dominant Negative Helix-Loop-Helix Protein	x						(216)
IDH1	Isocitrate dehydrogenase 1		x					(217)
IDH2	Isocitrate dehydrogenase 2		x					(217)
IKZF1	IKAROS family zinc finger 1	x	x					(218)
IKZF2	IKAROS family zinc finger 1		x					(219)
IKZF3	IKAROS family zinc finger 1		x					(219)
IL3	Interleukin 3			x				(202)
IL3RA	Interleukin 3 Receptor, Alpha (Low Affinity)	x						(220)

ITPR3	Inositol 1,4,5-Trisphosphate Receptor, Type 3				x			
KLF1	Kruppel-Like Factor 4 (Gut)			x				(221, 222)
LAPTM4B	Lysosomal Protein Transmembrane 4 Beta	x						(103)
LEF1	Lymphoid enhancer binding factor 1	x						(223)
LMO2	LIM domain only 2 (rhombotin-like 1)	x		x			x	(224)
LYL1	Lymphoblastic Leukemia Associated Hematopoiesis Regulator 1	x		x		x		(225)
MAP7	Microtubule associated protein 7	x						(103)
MECOM (EVI1)	MDS1 And EVI1 Complex Locus	x				x		(18)
MEIS1	Meis Homeobox 1	x		x		x		(226)
MLL	Myeloid/Lymphoid Or Mixed-Lineage Leukemia	x	x					(227)
MN1	Meningioma (Disrupted In Balanced Translocation) 1	x						(228)
MYB	MYB proto-oncogene, transcription factor					x		(229, 230)
MYBBP1A	MYB Binding Protein (P160) 1a				x			
MYC	v-myc avian myelocytomatosis viral oncogene homolog	x						(231)
NDFIP1	Nedd4 family interacting protein	x						(103)

	1							
NFATC2	Nuclear Factor Of Activated T-Cells, Cytoplasmic, Calcineurin-Dependent 2						x	(232)
NOTCH1	Notch homolog 1, translocation-associated (Drosophila)					x		(233)
NR4A3	Nuclear Receptor Subfamily 4, Group A, Member 3				x			(234)
PBX3	Pre B cell leukemia homeobox 3	x						(103)
PLA2G4A	Phospholipase A2 group IVA	x						(103)
PLOD3	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	x						(103)
PML	Promyelocytic leukemia						x	(235)
PPARG	Peroxisome proliferator activated receptor gamma					x		(44)
PRAME	Preferentially Expressed Antigen In Melanoma	x						(236)
PRDX2	Peroxiredoxin 2	x						(111)
PTEN	Phosphatase And Tensin Homolog	x						(237)
PTP4A3	protein tyrosine phosphatase type IVA, member 3	x						(103)
RAD21	RAD21 cohesin complex component		x					(178)
RARA	Retinoic acid receptor, alpha				x		x	(48)

RPLP0	Ribosomal Protein, Large, P0							
RUNX1	Runt-related transcription factor 1		x	x	x			(13)
SLC25A12	Solute carrier family 25 member 12	x						(103)
SLC2A5	Solute carrier family 2 member 5	x						(103)
SPI1 (PU.1)	Spleen focus forming virus (SFFV) proviral integration oncogene spi1		x	x	x		x	(46)
SRP72	Signal recognition particle 72kDa		x					(238)
SUZ12	Suppressor of Zeste 12		x					(239)
TAL1 (SCL)	T-cell acute lymphocytic leukemia 1			x	x	x	x	(13)
TERC	Telomerase RNA component		x					(238)
TERT	Telomerase reverse transcriptase		x					(238)
TET2	Ten-eleven translocation 2		x					(240)
TMEM159 (aka LOC57146)	Transmembrane protein 159	x						(103)
TP53	Tumor Protein P53	x						(241)
TRIM44	Tripartite motif containing 44	x						(103)
TRPS1	Transcriptional repressor GATA binding 1	x						(103)
UTX (KDM6A)	Ubiquitously Transcribed Tetratricopeptide Repeat, X Chromosome	x	x					(242)

VAV3	Vav guanine nucleotide exchange factor 3	x						(103)
VEGFA (aka VEGF)	Vascular endothelial growth factor A	x						(243)
WT1	Wilms Tumor 1	x	x					(220)
WT2 (H19)	Wilms tumor 2	x						(244)
ZBTB16 (PLZF)	Zinc finger and BTB domain containing 16						x	(49)
ZFPM1 (FOG1)	Zinc finger protein, multitype 1						x	(245)
ZFPM2 (FOG2)	Zinc finger protein, multitype 2						x	(246)

Appendix VII: Sequences of primers used for q-RTPCR using Fluidigm system in Chapter 4

Oligo Name	Sequence
ALS2CR8 F	TCACTTGTGCCATTCCTGTC
ALS2CR8 R	AACTTCAGTCATTGGGGGAA
ANGEL1 F	AAGGACAGACTCAGCCCAAC
ANGEL1 R	GACTTCCTGCTACGTTTCCG
ARL6IP5 F	ACGTCTTTATTGTGGGCTGC
ARL6IP5 R	GCTGCCATGATGATTTCCAT
AR F	AGTCAATGGGCAAAACATGG
AR R	TTGTGTCAAAAGCGAAATGG
ASXL1 F	GACCCACAGCTCTCCACATC
ASXL1 R	GAATCAGCCTTTTCACGCTC
B2M F	AATGTCGGATGGATGAAACC
B2M R	TCTCTCTTTCTGGCCTGGAG
BAALC-b F	AGGGCAGTCCATCTTCCAG
BAALC-b R	ACAGAATCCACCTGGCTCAC
BAP1 F	CGATCCATTTGAACAGGAAGA
BAP1 R	CTCGTGGAAGATTTCGGTGT
BMI1 F	CAGGTGGGGATTTAGCTCAG
BMI1 R	CTTTCATTGTCTTTTCCGCC
BSPRY F	GATAACCAGTTCCCTCGCTG
BSPRY R	GCAGTTACAGAGTGCTGCCA
BTBD3 F	CAGCAGCCAAGTCAATTTCA
BTBD3 R	ACGGAGAACTTGCAGAGGAC
C1RL F	ACCACAGAACTGGCTTGGAT
C1RL R	AGGCTCGTCTTCCAGGACTT
CD34 F	ATTTGAAAATGTTCCCTGGGT
CD34 R	TTTGCTTGCTGAGTTTGCTG
CEBPA F	TTCACATTGCACAAGGCACT
CEBPA R	GAGGGACCGGAGTTATGACA
CPT1A F	GCCTCGTATGTGAGGCAAAA
CPT1A R	TCATCAAGAAATGTCGCACG
CSF2-b F	GTCTCACTCCTGGACTGGCT
CSF2-b R	ACTACAAGCAGCACTGCCCT
CSF2RB F	ACACTCCAGGTCAGCAGGAA

CSF2RB R	CTCATTCCAACCAGACAGGC
DAPK1 F	ACTGGAGGCCGGTACTTTTC
DAPK1 R	GCAGGAAAACGTGGATGATT
DNMT3A F	ATAGATCCCGGTGTTGAGCC
DNMT3A R	ACCCAGCGCAGAAGCAG
EED F	TCTTCCATCTTGCCAGGTTT
EED R	ATGTTGATTGTGTGCGATGG
ERG F	TCTGTCTTAGCCAGGTGTGG
ERG R	CGCATTATGGCCAGCACTAT
ETFB F	CAGAAGGGGTTCATGGAGTG
ETFB R	GCTCGTAGCTGTCAAGAGGG
EVI2A F	TGTTGTCATCAGAAAGGCAAG
EVI2A R	CAGATTTTGACCAAGCATTTTG
EZH2-b F	CGCTTTTCTGTAGGCGATGT
EZH2-b R	CCGCTTATAAGTGTTGGGTGT
EZH2 F	CCCTTCTCAGATTTCTTCCCA
EZH2 R	GGACTCAGAAGGCAGTGGAG
FGFR1 F	GGCTGCCAAGACAGTGAAGT
FGFR1 R	GATGCTCCAGGTGGCATAAC
FLI1 F	CGCTGAGTCAAAGAGGGACT
FLI1 R	AATGTGTGGAATATTGGGGG
FLT3 F	TTGGGCATCATCATTTTCTG
FLT3 R	TGTGAGCAAAAGGGTCTTGA
FOSB F	GGTCCTGGCTGGTTGTGAT
FOSB R	TCTGTCTTCGGTGGACTCCT
GATA1 F	ACCAGAGCAGGATCCACAAA
GATA1 R	ATCACACTGAGCTTGCCACA
GATA2 F	GCCATAAGGTGGTGGTTGTC
GATA2 R	CTACCTGTGCAATGCCTGTG
GFI1B F	GAGCCATCTTGCTCTTCACC
GFI1B R	CGAGAGAGGCTTTGCAGTTC
HEATR6 F	GCTTTGAGCCTTCCAAGATG
HEATR6 R	AGCCCACAGTCAGTGTCCTT
HMBS F	GTACCCACGCGAATCACTCT
HMBS R	AGCCTACTTTCCAAGCGGAG
HOXA9 F	CAGTTCCAGGGTCTGGTGTT
HOXA9 R	AATGCTGAGAATGAGAGCGG
HPRT1 F	ACCCTTTCCAAATCCTCAGC

HPRT1 R	TCCTCCTCCTGAGCAGTCA
ID1 F	GACACAAGATGCGATCGTCC
ID1 R	AGTTGGAGCTGAACTCGGAA
IDH1 F	CTTTTGGGTTCCGTCACTTG
IDH1 R	GTCGTCATGCTTATGGGGAT
IDH2 F	TACGGGTCATCTCATCACCA
IDH2 R	ACCTCGCAAGAGCAGCC
IL3-b F	CACTTAAAGCAGCCACCTTTG
IL3-b R	CTGTTGAATGCCTCCAGGTT
IL3RA F	TCACGAAGACACAGACCAGG
IL3RA R	TGTATGAATTCTTGAGCGCC
ITPR3-b F	CGGAGAACAAGAAGGTGCAT
ITPR3-b R	GAAGCCGCTTGTTCACTGTC
KDM6A-b F	ATTCATAGCAGCGAACAGCC
KDM6A-b R	CTGGACAGCCGCCTCTT
KLF1 F	GGCTGGTCCTCAGACTTCAC
KLF1 R	CCGGACACACAGGATGACTT
LAPTM4B F	ATCAGCCAGGGCACTCAATA
LAPTM4B R	ACGCGGTTCTACTCCAACAG
LMO2-b F	GGCGCCTCTACTACAAACTGG
LMO2-b R	CTTTGTCTTTCACCCGCATT
LYL1 F	CTGCCTTCTCAGTCATGGTG
LYL1 R	ACCAGGCTGCAAGAACAGTG
MAP7 F	TTGCACTTTGTAGCTGTCGG
MAP7 R	ACCATGGCGGAGCTAGGAG
MECOM F	GCGCAATGTCTGCAACTACT
MECOM R	TGGAAGCTGGCTCAAGTACA
MEIS1 F	TCATCATCGTCACCTGTGCT
MEIS1 R	ACGGCATCTACTCGTTCAGG
MLL F	TGAAGGAGACCTTGTGGGAC
MLL R	CCGTGTTTGGGGAGAGC
MN1-b F	GTGCAGTGGACAGACAGGC
MN1-b R	GACGACGACAAGACGTTGG
MYBBP1A F	TCCCAGACTTTTCCTGATGC
MYBBP1A R	CTTCCAGCACCTTCTGCTCT
MYB F	GCAGGTTCCCAGGTACTGCT
MYB R	GCACCAGCATCAGAAGATGA
MYC F	CACCGAGTCGTAGTCGAGGT

MYC R	GCTGCTTAGACGCTGGATTT
NDFIP1 F	AGCCTGTTCAGGTTCTCCAG
NDFIP1 R	TCTGCTTCCCTGCTGCC
NFATC2 F	GGCTTGTTTTCCATGTAGCC
NFATC2 R	GGGCCCACTATGAGACAGAA
NOTCH1 F	ATAGTCTGCCACGCCTCTG
NOTCH1 R	AGTGTGAAGCGGCCAATG
NR4A3-b F	AAGCCTTAGCCTGCCTGTC
NR4A3-b R	GCCTGTCCCTTACTCTGGTG
NR4A3 F	CTGCTTGTGATCTTGTTGCAT
NR4A3 R	CTTAGCCTGCCTGTCAGCAC
PBX3 F	CCAGTCTCATTAGCTGGGGA
PBX3 R	TCTTCAGCGTCCTGTGTGAG
PLA2G4A F	TGCTCCACTATAATGTGCTGG
PLA2G4A R	GCATTGAGGAGCCTGAAGAT
PLOD3 F	ACACGGTTCCGATCAAACTT
PLOD3 R	CTACCTGGACCCAGGACTGA
PML F	GGAACATCCTCGGCAGTAGA
PML R	AGTCGGTGCGTGAGTTCC
PPARG F	TGCAGTGGGGATGTCTCATA
PPARG R	CCAACAGCTTCTCCTTCTCG
PRAME F	TGCCAGCTCCACAAGTCTC
PRAME R	GCTTCAAAATGGAACGAAGG
PRDX2 F	TGGGCTTAATCGTGTCACTG
PRDX2 R	TTAATGATTTGCCTGTGGGA
PTEN F	TTGGCGGTGTCATAATGTCT
PTEN R	GCAGAAAGACTTGAAGGCGTA
PTP4A3 F	CTACAAACACATGCGCTTCC
PTP4A3 R	AGCCCCGTACTTCTTCAGGT
RAD21 F	CAAATTTTGGCCAGAGGC
RAD21 R	ACTCCCAGCGGAGAGCA
RARA F	AGGGCTGGGCACTATCTCTT
RARA R	CCTATGCTGGGTGGACTCTC
RPLP0 F	GGATCTGCTGCATCTGCTTG
RPLP0 R	GCGACCTGGAAGTCCAACTA
RUNX1 F	GGCATCGTGGACGTCTCTA
RUNX1 R	CGATGGCTTCAGACAGCATA
SFPQ F	CCCATTCCTCTAGGACCCTG

SFPQ R	TGAAGCTAATCCTGGCGTTC
SLC25A12 F	CAGTCCAAGATAGCGCTGAA
SLC25A12 R	CAGACAACTAAGCGAGGGGA
SLC2A5 F	TGACAGCAGCCACGTTGTA
SLC2A5 R	GCAACAGGATCAGAGCATGA
SON F	CAACATCGAGCAGATTTTTAGGT
SON R	TGGTTTCCTTCAATGGGTGTA
SPI1-b F	GAAGACCTGGTGCCCTATGA
SPI1-b R	GGGGTGGAAGTCCCAGTAAT
SRP72 F	GCAGCCATCTTTCTGGATCT
SRP72 R	CAAAGGAACAAGGACAGGGA
SUZ12-b F	GCAGGACTTCCAGGGTAACA
SUZ12-b R	TGAAAGGAGAGCAAGAATCTCA
TAL1 F	GATGTGTGGGGATCAGCTTG
TAL1 R	GGAGACCTTCCCCCTATGAG
TERC F	ACCCTAACTGAGAAGGGCGTA
TERC R	GCTCTAGAATGAACGGTGGAA
TERT F	CCGCCTGAGCTGTACTTTGT
TERT R	CAGTACGTGTTCTGGGGTTTG
TET2 F	GGCTACAAAGCTCCAGAATGG
TET2 R	AAGAGTGCCACTTGGTGTCTC
TMEM159 F	GGCTGTCCAAGTACTGACCC
TMEM159 R	GCAGGAACTGCAGAAGAAGC
TP53 F	GCTCGACGCTAGGATCTGAC
TP53 R	GCTTTCCACGACGGTGAC
TRIM44 F	TGATTTTCCACTGGGGGTAG
TRIM44 R	TGAATCAGCTGAGCCAAAGG
TRPS1 F	ACATATCCGCCATTTGCATT
TRPS1 R	AAGGATGAATCCCAGTCCCT
VAV3 F	TGTGGCCAATGCTATAGGTG
VAV3 R	TGAACTTTTCGAGGCATTTG
VEGFA F	AGCTGCGCTGATAGACATCC
VEGFA R	CTACCTCCACCATGCCAAGT
WT1 F	TTGTGTGGTTATCGCTCTCG
WT1 R	CAAATGACATCCCAGCTTGA
WT2-H19 F	CTCAGCGTTCGGGCTGGAG
WT2-H19 R	ACCTGGCGTCTTGGCCTTC
ZBTB16 F	TTCTCAGCCGCAAACTATCC

ZBTB16 R	ATAACGAGGCTGTGGAGCAG
ZFPM1-b F	TGAAGAAGGAGCCAGCAGAG
ZFPM1-b R	ACAGTCCTTGCAGGGGAAGA
ZFPM2 F	TCTTCACCCTCAGAGATGGC
ZFPM2 R	CTTGGCAAGGAGTGGAAGAC

Appendix VIII. Ranking of genes based on random forest analysis for different survival outcomes

Rank	OS	DSS	DFS	EFS
1	***BSPRY***	***BSPRY***	***LAPTM4B***	***LAPTM4B***
2	***LAPTM4B***	*RARA*	***BSPRY***	***BSPRY***
3	***IDH1***	***ERG***	*PRAME*	***NR4A3***
4	***NR4A3***	*LAPTM4B*	*RARA*	*ETFB*
5	*ETFB*	*HEATR6*	*ETFB*	*DAPK1*
6	***ERG***	***IDH1***	***NR4A3***	*PRAME*
7	*SPI1*	*LMO2*	*PLA2G4A*	***IDH1***
8	*DAPK1*	***NR4A3***	***ERG***	***ERG***
9	*HEATR6*	*TP53*	***IDH1***	*RARA*
10	*LMO2*	*FLI1*	*LMO2*	*SPI1*
11	*BTBD3*	*ETFB*	*HOXA9.1*	*LMO2*
12	*RARA*	*SPI1*	*TET2*	*CSF2RB*
13	*CSF2RB*	*BTBD3*	*WT1*	*BTBD3*
14	*TET2*	*EZH2*	*BTBD3*	*HEATR6*
15	*PRAME*	*SLC2A5*	*DNMT3A*	*TET2*
16	*TP53*	*CTCF*	*CTCF*	*HOXA9.1*
17	*IKZF1*	*WT1*	*SPI1*	*HMBS*
18	*SLC2A5*	*TERC*	*HEATR6*	*RAD21*
19	*FLI1*	*TET2*	*EZH2*	*MYBBP1A*
20	*MYBBP1A*	*PBX3*	*TERC*	*PPARG*
21	*DNMT3A*	*IKZF3*	*PBX3*	*PBX3*
22	*WT1*	*ANGEL1*	*TERT*	*SUZ12*
23	*TMEM159*	*HOXA9.1*	*FLI1*	*WT1*
24	*ASXL1*	*MYB*	*RAD21*	*MYB*
25	*SUZ12*	*RAD21*	*TP53*	*IKZF1*
26	*EZH2*	*C1RL*	*ITPR3*	*TERC*
27	*HOXA9.1*	*PTP4A3*	*PPARG*	*ASXL1*
28	*LEF1*	*CSF2RB*	*IDH2*	*B2M*
29	*EVI2A*	*PRAME*	*TMEM159*	*DNMT3A*
30	*HMBS*	*VAV3*	*MYC*	*GATA1*
31	*IKZF3*	*DNMT3A*	*HOXA9*	*CTCF*
32	*MLL*	*CSF2*	*PLOD3*	*PLA2G4A*
33	*MYB*	*MEIS1*	*CD34*	*EVI2A*
34	*KLF1*	*PLOD3*	*C1RL*	*VAV3*
35	*NFATC2*	*PLA2G4A*	*MN1*	*ALS2CR8*
36	*HOXA9*	*TRIM44*	*BAP1*	*FLT3*

37	*CTCF*	*BAP1*	*MYB*	*BAP1*
38	*PTEN*	*HOXA9*	*SUZ12*	*TERT*
39	*CPT1A*	*DAPK1*	*CSF2RB*	*MYC*
40	*MEIS1*	*B2M*	*LEF1*	*PRDX2*
41	*SRP72*	*ITPR3*	*MYBBP1A*	*RPLP0*
42	*PTP4A3*	*LEF1*	*CSF2*	*LYL1*
43	*UTX*	*IKZF1*	*IKZF3*	*CSF2*
44	*MAP7*	*NFATC2*	*ASXL1*	*HOXA9*
45	*ZBTB16*	*SRP72*	*ALS2CR8*	*IKZF3*
46	*RUNX1*	*IKZF2*	*SRP72*	*ANGEL1*
47	*ANGEL1*	*MYBBP1A*	*SLC25A12*	*TAL1*
48	*B2M*	*SUZ12*	*AR*	*CD34*
49	*BAP1*	*TRPS1*	*TRPS1*	*RUNX1*
50	*IKZF2*	*MN1*	*PTP4A3*	*TRIM44*
51	*C1RL*	*TERT*	*EVI2A*	*MECOM*
52	*RAD21*	*SLC25A12*	*UTX*	*EZH2*
53	*GATA1*	*UTX*	*FLT3*	*SRP72*
54	*PRDX2*	*PTEN*	*IKZF2*	*IDH2*
55	*TRIM44*	*ZBTB16*	*ZBTB16*	*TP53*
56	*TERT*	*ASXL1*	*PRDX2*	*SLC25A12*
57	*CD34*	*PRDX2*	*TAL1*	*PML*
58	*ITPR3*	*MLL*	*NFATC2*	*ZBTB16*
59	*TERC*	*FGFR1*	*ANGEL1*	*MN1*
60	*VAV3*	*CD34*	*IKZF1*	*FLI1*
61	*PBX3*	*HMBS*	*PML*	*MAP7*
62	*MN1*	*KLF1*	*B2M*	*LEF1*
63	*SLC25A12*	*RPLP0*	*VAV3*	*TRPS1*
64	*PPARG*	*EVI2A*	*PTEN*	*MEIS1*
65	*PLA2G4A*	*HPRT1*	*MLL*	*PTP4A3*
66	*MECOM*	*MYC*	*HMBS*	*KLF1*
67	*TAL1*	*GATA1*	*DAPK1*	*ITPR3*
68	*ALS2CR8*	*MAP7*	*HPRT1*	*AR*
69	*CSF2*	*BMI1*	*KLF1*	*MLL*
70	*FLT3*	*FLT3*	*CEBPA*	*TMEM159*
71	*IDH2*	*ALS2CR8*	*RUNX1*	*BMI1*
72	*BMI1*	*MECOM*	*TRIM44*	*HPRT1*
73	*CEBPA*	*TAL1*	*MECOM*	*UTX*
74	*TRPS1*	*PML*	*GATA1*	*SLC2A5*
75	*RPLP0*	*CPT1A*	*MEIS1*	*PTEN*
76	*AR*	*IDH2*	*SLC2A5*	*NFATC2*
77	*PML*	*AR*	*LYL1*	*C1RL*
78	*HPRT1*	*VEGFA*	*GFI1B*	*CEBPA*

79	*MYC*	*RUNX1*	*BMI1*	*CPT1A*
80	*LYL1*	*TMEM159*	*FGFR1*	*PLOD3*
81	*PLOD3*	*LYL1*	*RPLP0*	*FGFR1*
82	*GFI1B*	*PPARG*	*VEGFA*	*GFI1B*
83	*VEGFA*	*GATA2*	*CPT1A*	*GATA2*
84	*FGFR1*	*CEBPA*	*MAP7*	*IKZF2*
85	*GATA2*	*GFI1B*	*GATA2*	*VEGFA*

Appendix IX. Endpoints and definitions of survival values

Overall Survival: From date of diagnosis to date of transplant (censored), OR date of death (not censored), OR last date of contact (censored).
Disease Specific Survival: From date of diagnosis to date of transplant (censored), OR date of death from AML (not censored), OR data of death from unrelated causes (censored), OR last date of contact (censored).
Disease Free Survival: From date of diagnosis to date of relapse before transplant (not censored), OR date of transplant (censored), OR date of relapse after transplant (not censored), OR date of death from AML (not censored), OR date of death from unrelated causes (censored), OR last date of contact (censored).
Event Free Survival: From date of diagnosis to date of relapse before transplant (not censored), OR date of transplant (censored), OR date of death (not censored), OR last date of contact (censored).

Appendix X. Supplementary Methods and Figures for Chapter 6.

Self-reverting mutations partially correct the blood phenotype in a Diamond Blackfan Anemia patient

Parvathy Venugopal[1-3], Sarah Moore[1], David M Lawrence[3,4], Amee J George[5-7], Ross D Hannan[5-10], Sarah CE Bray[2,13], Luen Bik To[11,13], Richard J D'Andrea[2,11,12], Jinghua Feng[4,12] , Amanda Tirimacco[1], Alexandra L Yeoman[1], Chun Chun Young[1] , Miriam Fine[14], Andreas W Schreiber[3,4], Christopher N Hahn[1,2,13], Christopher Barnett[13,14], Ben Saxon[12, 15] and Hamish S Scott[1-3,12,13 #]

This file contains Materials and Methods, Supplementary Figures as well as the Supplementary References.

Methods:

Ethics: Experiments were carried out under ethical approval by the Peter MacCallum Cancer Centre Ethics committee (HREC 13-185) and the Australian Familial Haematological Cancer Study (REC1542/12/2015) approved by the Women's and Children's Hospital Human Research Ethics Committee in accordance with the Declaration of Helsinki.

Generation of lymphoblastoid cell lines (LCLs): Epstein Barr virus-transformed lymphoblastoid cell-lines (LCLs) were generated from PBMNCs isolated from the trio and two siblings, and three healthy donors. LCLs were cultured in RPMI-1640 with HEPES (Gibco Invitrogen) supplemented with 10% FBS (Sigma Aldrich) at 37^{o}C with 5% CO_2.

Analysis of SBDS protein expression: LCLs were washed once with PBS and pelleted at 500 x g for 5 minutes. Cell pellets were lysed in SDS lysis buffer (20 mM HEPES pH 7.9, 0.5 mM EDTA and 2% (w/v) SDS) and protein concentration quantified using the BioRad D_C protein assay as per manufacturers' instructions. Protein samples (50 µg) were then mixed with Laemmli sample buffer (1) containing 8% β-mercaptoethanol, and heated at 95^{o}C for 5 minutes. Protein samples were then electrophoresed on Novex 4-20% Tris Glycine SDS-PAGE gels (ThermoFisher Scientific), and transferred onto PVDF membrane (Immobilon-P, Merck Millipore)

using the BioRad Trans-Blot semi-dry transfer cell using standard settings. Membranes were then blocked in 5% (w/v) skim milk (Diploma Brand, Fonterra Food Services) in Tris buffered saline (TBS) pH 7.6 containing 0.05% (v/v) Tween 20 (TBST). Membranes were then subsequently immunoblotted with primary and secondary HRP-conjugated antibodies (all prepared in 5% skim milk in TBST), and visualised using enhanced chemiluminescence (Western Lighting Plus ECL Kit, Perkin Elmer) on Hyperfilm (GE Life Sciences). SBDS antibody [EPR7820] (ab128946) was purchased from Abcam and the β-actin antibody was purchased from MP Biosciences (691002). Secondary antibodies (goat-α-mouse and goat-α-rabbit were purchased from BioRad (172-1011 and 170-6515, respectively).

Analysis of SBDS mRNA expression (including RNA isolation and cDNA synthesis): LCLs were washed once with PBS and pelleted at 500 x g for 5 minutes and RNA extracted using the Qiagen miRNeasy mini kit as per manufacturer's instructions. Isolated RNA (2 µg) was then subjected to RQ1 DNase treatment (Promega) for 30 minutes at 37^{o}C, followed by reverse transcription using the Superscript III kit (ThermoFisher Scientific) with random primers (Promega). cDNA was then stored at -20^{o}C prior to analysis. Taqman Fast Advanced Master Mix (#4444557) and Taqman primer/probe sets (SBDS, Hs04188846_m1; B2M, Hs00984230_m1) were utilised for real-time PCR quantitation of SBDS expression (available from Applied Biosystems, ThermoFisher Scientific). cDNA isolated as described above, was assayed using the aforementioned reagents on the Applied Biosystems StepOnePlus real-time PCR instrument as per the manufacturers' instructions using the instrument default cycling conditions. Data was analysed using the 7000 SDS 1.1 RQ Software (Applied Biosystems) where relative quantification of gene expression was performed (normalized to B2M expression).

Targeted Next Generation Sequencing (NGS): Trio gDNA was subjected to targeted next generation sequencing (NGS) of 94 genes and 284 SNPs associated with predisposition to cancer on the TruSight Cancer Panel (Illumina). gDNA from the proband (peripheral blood {PB}, hair) was also analysed on the 54 gene TruSight Myeloid panel. The MiSeq Reporter package was used to call variants and annotation was provided through our ACRF Cancer Genomics Facility custom pipeline, which takes into consideration pathogenicity/oncogenicity predictions (CADD>10,

Polyphen 2, SIFT, Mutation Taster, GERP > 2, COSMIC parameters including specific-mutation and gene frequency), population minor allele frequencies (1000 GP, ESP, ExAC), OMIM, Gene Ontology and various parameters designed to filter out systematic errors. Variants detected in PB and hair were analysed with VariantGrid (in-house analysis software) to identify somatic variants.

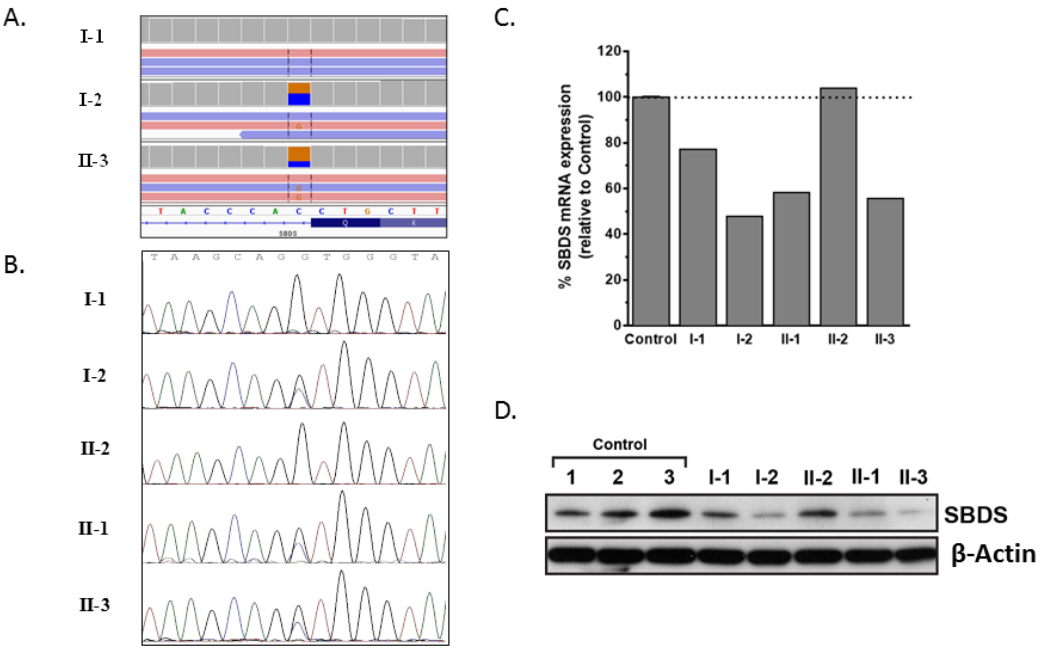
Whole Genome Sequencing: Genomic DNA samples from PB of the trio (proband, father and mother) were subjected to whole genome sequencing (WGS) by Illumina (San Diego, California, USA) with >30x coverage for more than 90% of the genome. Sequencing, alignment to the hg19 reference genome and variant calling were performed by Illumina using the Isaac software(2). The single read spanning the 184kb deletion was only detected after re-mapping reads with BWA-MEM(3). Variant annotation and filtration was also carried out by Illumina. Based on the reported family history, variants were evaluated based on autosomal recessive inheritance, *de novo* inheritance, x-linked inheritance, as well as dominant inheritance (based on literature reports of reduced penetrance for DBA). All variants meeting any of the above inheritance patterns, and found in any gene associated with DBA, short stature, anemia, or ribosomal protein were evaluated. Clinical interpretation was performed using the American College of Medical Genetics and Genomics guidelines for evaluation and classification of genetic variant information. In addition to this, clinical interpretation was performed for all single nucleotide variants, insertions up to 3 nucleotides in length, and deletions up to 11 nucleotides in length, in a subset of genes recommended by the American College of Medical Genetics and Genomics (ACMG). All calls within the subset of genes located in the Gene list appendix were evaluated for evidence of clinical importance including: allele frequency in population studies (dbSNP, 1000 Genomes, *etc*.), evidence in the scientific literature for likely causation of the condition, and consideration of the likely biological implications of the variant based on its expected characteristics.

SNP array: gDNA from the PB of the trio, and hair, bone marrow and lymphoblastoid cell lines (LCLs) of the proband were also analysed by high density CytoSNP 850K BeadArray (Illumina).

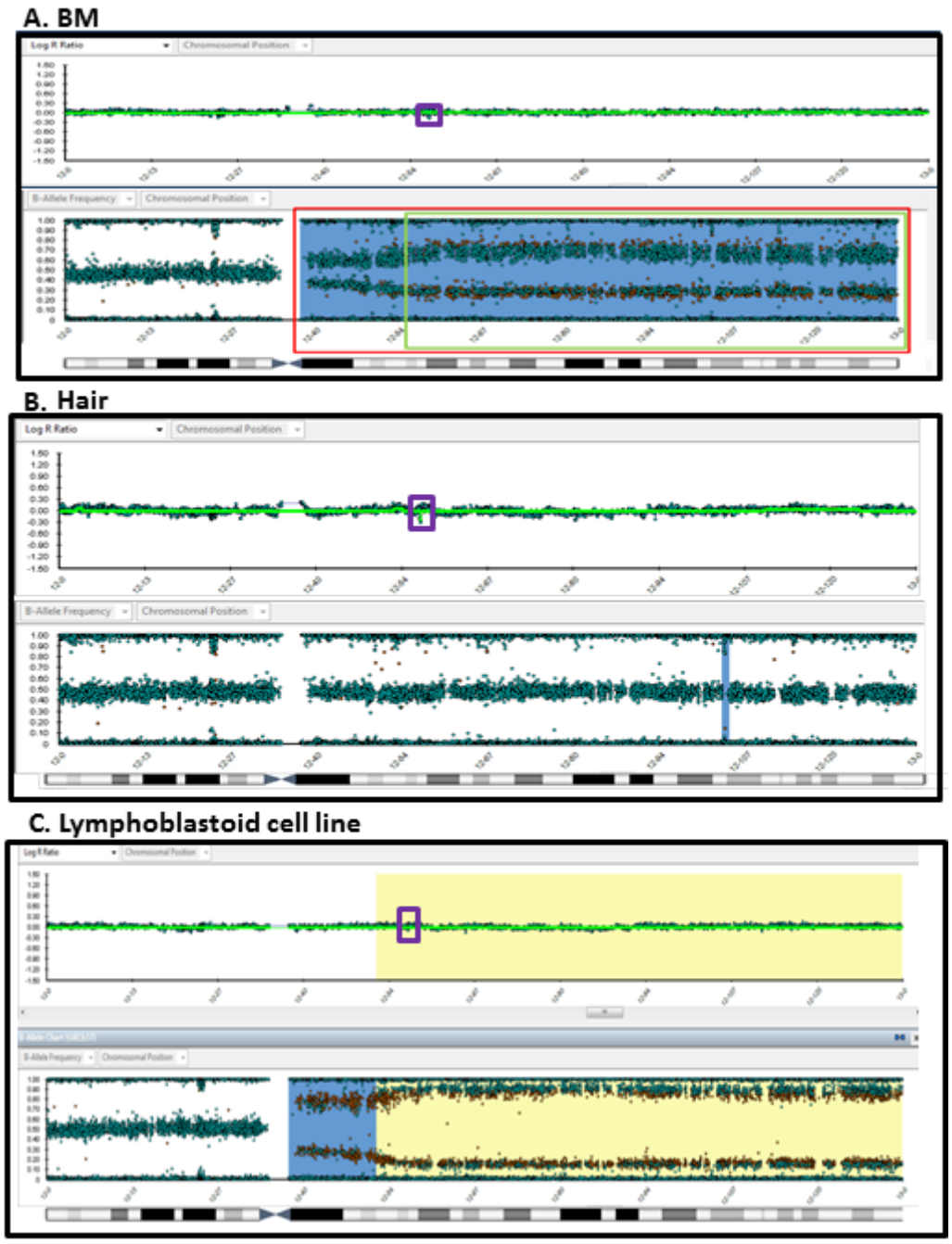
Erythroid colony assays: Erythroid burst forming units from PBMNCs were grown in MethoCult (#H4230, Stem Cell Technologies) in the presence of SCF (50 ng/ml),

IL-3 (20 ng/ml) and EPO (2 U/ml) for 14 days. Colonies were scored and picked. Whole genome amplification was performed on colony genomic DNA using the GenomePlex Complete WGA Kit (WGA2, Sigma Aldrich). Sanger sequencing was performed across informative selected SNPs on chromosome 12, to genotype the colonies for cnLOH events.

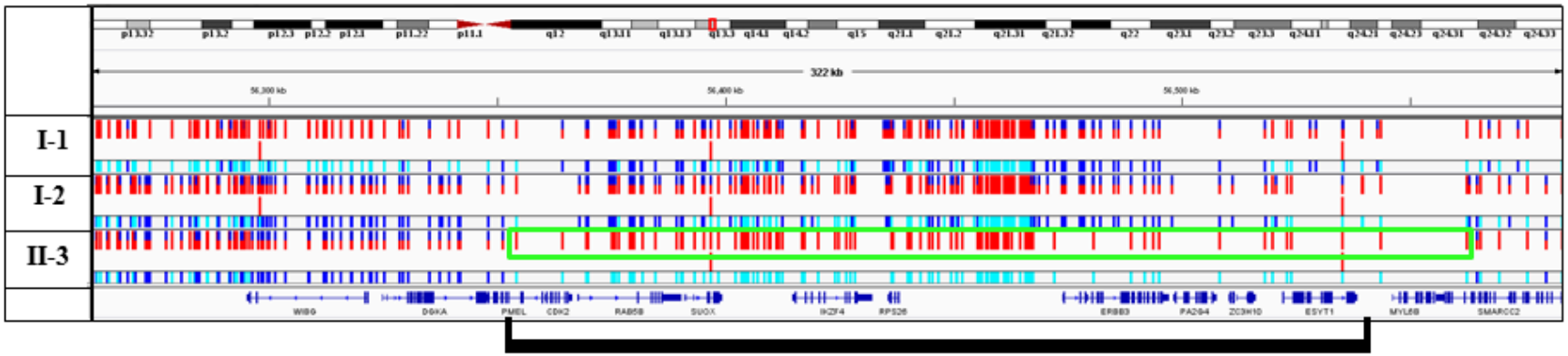
Supplementary Figure Legends:



Supplementary Figure 1. Mutation in splice donor site in SBDS. WGS identified a c.258+1G>C splice donor site mutation in the proband and mother (A). This mutation was confirmed by Sanger sequencing in the proband, mother and brother (B). Quantitative RT-PCR (C) and western blot analysis (D) reveal lower expression of SBDS in individuals with the mutated allele compared to normal controls.



Supplementary Figure 2. Deletion in chromosome 12q found in BM and hair of the proband but regions of copy neutral loss of heterozygosity (cnLOH) restricted to BM and LCL. (A) SNP microarray confirms 184 kb deletion encompassing 11 genes on chr 12q in BM (purple box) and reveals cnLOH comprising two regions of chr 12q (denoted by green and red boxes). (B) 184 kb deletion confirmed to be germline by SNP microarray on DNA from hair. (C) cnLOH events confirmed in larger proportions in LCL when compared to BM.

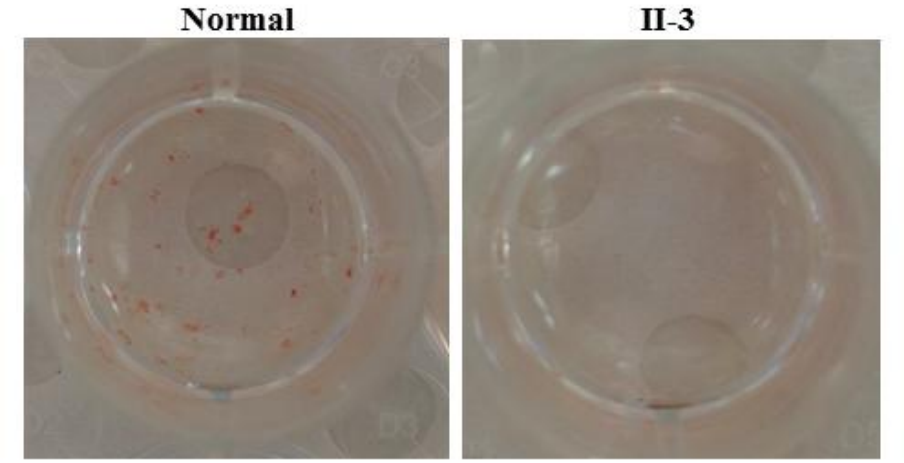


Supplementary Figure 3. Loss of heterozygosity within 184 kb deletion in proband. Region of loss of heterozygosity (green box) encompasses 184 kb deletion (black bar) in the proband (II-3). The parents are heterozygous across this region. Alternative alleles are depicted by blue and red bars.

```
AluY        GAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCAGAGCTTGCAGTGAGCCGAGATT
Sanger      GAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCAGAGCTTGCAGTGAGCCGAGATC
AluSc       GAGGCTGAAACAGGAGAGTCGCTTGAACCAGGGAGGCGGAGGTTGCAGTGAGCCAAGATG
            ********  ******* * ** ****** ******* *** ************ ****

AluY        GCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAGAAAAAAAA
Sanger      GCGCCACTGTACTCCAGCCTGGGCGACACAGCGAGACTCCATCTCAAAAAAAAAAAA---
AluSc       GCGCCACTGCACTCTAGCCTGGGCGACACAGCGAAACTCCATCTCAAAAAAAAAAAA---
            ********* **** ************ ***** ***** ********** *****
```

Supplementary Figure 4. Breakpoints of the 184 kb deletion lie within Alu repeats with high sequence similarity. 184 kb deletion is likely the result of homologous recombination event between Alu repeats – AluY (56354900-56355009) and AluSc8 (56539460-56539754) which show 88% sequence identity. The breakpoint has been narrowed down to a region of 56 bp using Sanger sequencing (underlined). Regions that are identical to AluY (green) and AluSc8 (blue) are highlighted.



Supplementary Figure 5. Proband exhibits significantly reduced erythroid colony forming ability *in vitro*. PBMNCs from the proband formed very few erythroid colonies (5 colonies, 3 colonies) in comparison with a normal (58 colonies, 46 colonies) when tested in a methylcellulose based assay to quantify erythroid burst forming units. The colonies that did form were much smaller than normal colonies.

Supplementary Table 1. Hematological parameters of the proband. Blood parameters of the proband from age 9 months to present.

Supplementary Table 2. Genes affected by germline mutations or deletions. Functions and phenotypes associated with mutated/deleted genes(4-12).

Supplementary Table 3. Allelic imbalance in the proband across chr 12q. Analysis of reads across bases where proband is heterozygous in the targeted NGS (TruSight Cancer) panel, confirmed allelic imbalance with an over-representation of the paternal allele.

Supplementary References

1. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227(5259):680-5.
2. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics. 2013;29(16):2041-3.
3. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.
4. Ruggero D, Shimamura A. Marrow failure: a window into ribosome biology. Blood. 2014;124(18):2784-92.
5. Hellstrom AR, Watt B, Fard SS, Tenza D, Mannstrom P, Narfstrom K, et al. Inactivation of Pmel alters melanosome shape but has only a subtle effect on visible pigmentation. PLoS genetics. 2011;7(9):e1002285.
6. Narkis G, Ofir R, Manor E, Landau D, Elbedour K, Birk OS. Lethal congenital contractural syndrome type 2 (LCCS2) is caused by a mutation in ERBB3 (Her3), a modulator of the phosphatidylinositol-3-kinase/Akt pathway. American journal of human genetics. 2007;81(3):589-95.
7. Berthet C, Aleem E, Coppola V, Tessarollo L, Kaldis P. Cdk2 knockout mice are viable. Current biology : CB. 2003;13(20):1775-85.
8. Chiariello M, Bruni CB, Bucci C. The small GTPases Rab5a, Rab5b and Rab5c are differentially phosphorylated in vitro. FEBS letters. 1999;453(1-2):20-4.

9. Zhang SJ, Ma LY, Huang QH, Li G, Gu BW, Gao XD, et al. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(6):2076-81.

10. Steinberg KK, Relling MV, Gallagher ML, Greene CN, Rubin CS, French D, et al. Genetic studies of a cluster of acute lymphoblastic leukemia cases in Churchill County, Nevada. Environmental health perspectives. 2007;115(1):158-64.

11. Fu W, Ergun A, Lu T, Hill JA, Haxhinasto S, Fassett MS, et al. A multiply redundant genetic switch 'locks in' the transcriptional signature of regulatory T cells. Nature immunology. 2012;13(10):972-80.

12. Wang S, Huang J, He J, Wang A, Xu S, Huang SF, et al. RPL41, a small ribosomal peptide deregulated in tumors, is essential for mitosis and centrosome integrity. Neoplasia (New York, NY). 2010;12(3):284-93.

References

1. Bryder D, Rossi DJ, Weissman IL. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. The American journal of pathology. 2006;169(2):338-46.
2. Cantor AB, Iwasaki H, Arinobu Y, Moran TB, Shigematsu H, Sullivan MR, et al. Antagonism of FOG-1 and GATA factors in fate choice for the mast cell lineage. The Journal of experimental medicine. 2008;205(3):611-24.
3. Akala OO, Clarke MF. Hematopoietic stem cell self-renewal. Current opinion in genetics & development. 2006;16(5):496-501.
4. Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nature immunology. 2008;9(2):129-36.
5. Robb L, Lyons I, Li R, Hartley L, Kontgen F, Harvey RP, et al. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. Proceedings of the National Academy of Sciences of the United States of America. 1995;92(15):7075-9.
6. Tsai FY, Keller G, Kuo FC, Weiss M, Chen J, Rosenblatt M, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature. 1994;371(6494):221-6.
7. Zhou Y, Lim KC, Onodera K, Takahashi S, Ohta J, Minegishi N, et al. Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development. The EMBO journal. 1998;17(22):6689-700.
8. Yamada Y, Warren AJ, Dobson C, Forster A, Pannell R, Rabbitts TH. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(7):3890-5.
9. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell. 1996;84(2):321-30.
10. Giampaolo A, Sterpetti P, Bulgarini D, Samoggia P, Pelosi E, Valtieri M, et al. Key functional role and lineage-specific expression of selected HOXB genes in purified hematopoietic progenitor differentiation. Blood. 1994;84(11):3637-47.
11. Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, Graf T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity. 2006;25(5):731-44.
12. Arinobu Y, Mizuno S, Chong Y, Shigematsu H, Iino T, Iwasaki H, et al. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell stem cell. 2007;1(4):416-27.
13. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell stem cell. 2010;7(4):532-44.
14. Moignard V, Macaulay IC, Swiers G, Buettner F, Schutte J, Calero-Nieto FJ, et al. Characterization of transcriptional networks in blood stem and progenitor cells using

high-throughput single-cell gene expression analysis. Nature cell biology. 2013;15(4):363-72.
15. Meyers CA, Albitar M, Estey E. Cognitive impairment, fatigue, and cytokine levels in patients with acute myelogenous leukemia or myelodysplastic syndrome. Cancer. 2005;104(4):788-93.
16. Slovak ML, Kopecky KJ, Cassileth PA, Harrington DH, Theil KS, Mohamed A, et al. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. Blood. 2000;96(13):4075-83.
17. Lugthart S, Groschel S, Beverloo HB, Kayser S, Valk PJ, van Zelderen-Bhola SL, et al. Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(24):3890-8.
18. Lugthart S, van Drunen E, van Norden Y, van Hoven A, Erpelinck CA, Valk PJ, et al. High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. Blood. 2008;111(8):4329-37.
19. Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nature genetics. 2011;43(10):1012-7.
20. Yamazaki H, Suzuki M, Otsuki A, Shimizu R, Bresnick EH, Engel JD, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. Cancer cell. 2014;25(4):415-27.
21. Groschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BA, Erpelinck C, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell. 2014;157(2):369-81.
22. Paschka P, Schlenk RF, Gaidzik VI, Herzig JK, Aulitzky T, Bullinger L, et al. ASXL1 mutations in younger adult patients with acute myeloid leukemia: a study by the German-Austrian Acute Myeloid Leukemia Study Group. Haematologica. 2015;100(3):324-30.
23. Vardiman JW. 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):16-20.
24. Corey SJ, Minden MD, Barber DL, Kantarjian H, Wang JC, Schimmer AD. Myelodysplastic syndromes: the complexity of stem-cell diseases. Nature reviews Cancer. 2007;7(2):118-29.
25. Vardiman J. The classification of MDS: from FAB to WHO and beyond. Leukemia research. 2012;36(12):1453-8.
26. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al. Proposals for the classification of the myelodysplastic syndromes. British journal of haematology. 1982;51(2):189-99.
27. Bennett JM, Brunning RD, Vardiman JW. Myelodysplastic syndromes: from French-American-British to World Health Organization: a commentary. Blood. 2002;99(8):3074-5.
28. Videbaek A. Familial Leukaemia. Acta Medica Scandinavica. 1947;CXXVII(I-II).

29. Rosenberg PS, Greene MH, Alter BP. Cancer incidence in persons with Fanconi anemia. Blood. 2003;101(3):822-6.
30. German J. Bloom's syndrome. XX. The first 100 cancers. Cancer genetics and cytogenetics. 1997;93(1):100-6.
31. Bader JL, Miller RW. Neurofibromatosis and childhood leukemia. The Journal of pediatrics. 1978;92(6):925-9.
32. Socie G, Henry-Amar M, Bacigalupo A, Hows J, Tichelli A, Ljungman P, et al. Malignant tumors occurring after treatment of aplastic anemia. European Bone Marrow Transplantation-Severe Aplastic Anaemia Working Party. The New England journal of medicine. 1993;329(16):1152-7.
33. Xue Y, Zhang R, Guo Y, Gu J, Lin B. Acquired amegakaryocytic thrombocytopenic purpura with a Philadelphia chromosome. Cancer genetics and cytogenetics. 1993;69(1):51-6.
34. Geissler D, Thaler J, Konwalinka G, Peschel C. Progressive preleukemia presenting amegakaryocytic thrombocytopenic purpura: association of the 5q- syndrome with a decreased megakaryocytic colony formation and a defective production of Meg-CSF. Leukemia research. 1987;11(8):731-7.
35. Owen CJ, Toze CL, Koochin A, Forrest DL, Smith CA, Stevens JM, et al. Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy. Blood. 2008;112(12):4639-45.
36. Ostergaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). Nature genetics. 2011;43(10):929-31.
37. Kazenwadel J, Secker GA, Liu YJ, Rosenfeld JA, Wildin RS, Cuellar-Rodriguez 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):1283-91.
38. Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. Blood. 2011;118(10):2653-5.
39. Tawana K, Fitzgibbon J. Inherited DDX41 mutations: 11 genes and counting. Blood. 2016;127(8):960-1.
40. Dokal I, Vulliamy T. Inherited bone marrow failure syndromes. Haematologica. 2010;95(8):1236-40.
41. Ruggero D, Shimamura A. Marrow failure: a window into ribosome biology. Blood. 2014;124(18):2784-92.
42. Talwalkar SS, Yin CC, Naeem RC, Hicks MJ, Strong LC, Abruzzo LV. Myelodysplastic syndromes arising in patients with germline TP53 mutation and Li-Fraumeni syndrome. Archives of pathology & laboratory medicine. 2010;134(7):1010-5.
43. Pan X, Minegishi N, Harigae H, Yamagiwa H, Minegishi M, Akine Y, et al. Identification of human GATA-2 gene distal IS exon and its expression in hematopoietic stem cell fractions. Journal of biochemistry. 2000;127(1):105-12.
44. Vicente C, Conchillo A, Garcia-Sanchez MA, Odero MD. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Critical reviews in oncology/hematology. 2012;82(1):1-17.

45. Chang AN, Cantor AB, Fujiwara Y, Lodish MB, Droho S, Crispino JD, et al. GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(14):9237-42.
46. Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, Radomska HS, et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(15):8705-10.
47. Ozawa Y, Towatari M, Tsuzuki S, Hayakawa F, Maeda T, Miyata Y, et al. Histone deacetylase 3 associates with and represses the transcription factor GATA-2. Blood. 2001;98(7):2116-23.
48. Tsuzuki S, Kitajima K, Nakano T, Glasow A, Zelent A, Enver T. Cross talk between retinoic acid signaling and transcription factor GATA-2. Molecular and cellular biology. 2004;24(15):6824-36.
49. Tsuzuki S, Enver T. 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):3404-10.
50. Chun TH, Itoh H, Subramanian L, Iniguez-Lluhi JA, Nakao K. Modification of GATA-2 transcriptional activity in endothelial cells by the SUMO E3 ligase PIASy. Circulation research. 2003;92(11):1201-8.
51. Rodrigues NP, Janzen V, Forkert R, Dombkowski DM, Boyd AS, Orkin SH, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood. 2005;106(2):477-84.
52. Rodrigues NP, Boyd AS, Fugazza C, May GE, Guo Y, Tipping AJ, et al. GATA-2 regulates granulocyte-macrophage progenitor cell function. Blood. 2008;112(13):4862-73.
53. Persons DA, Allay JA, Allay ER, Ashmun RA, Orlic D, Jane SM, et al. Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. Blood. 1999;93(2):488-99.
54. Tsai FY, Orkin SH. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood. 1997;89(10):3636-43.
55. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, Keles S. GATA switches as developmental drivers. The Journal of biological chemistry. 2010;285(41):31087-93.
56. Shimamoto T, Ohyashiki K, Ohyashiki JH, Kawakubo K, Fujimura T, Iwama H, et al. The expression pattern of erythrocyte/megakaryocyte-related transcription factors GATA-1 and the stem cell leukemia gene correlates with hematopoietic differentiation and is associated with outcome of acute myeloid leukemia. Blood. 1995;86(8):3173-80.
57. Ayala RM, Martinez-Lopez J, Albizua E, Diez A, Gilsanz F. Clinical significance of Gata-1, Gata-2, EKLF, and c-MPL expression in acute myeloid leukemia. American journal of hematology. 2009;84(2):79-86.
58. Luesink M, Hollink IH, van der Velden VH, Knops RH, Boezeman JB, de Haas V, et al. High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. Blood. 2012;120(10):2064-75.
59. Wieser R, Volz A, Vinatzer U, Gardiner K, Jager U, Mitterbauer M, et al. Transcription factor GATA-2 gene is located near 3q21 breakpoints in myeloid leukemia. Biochemical and biophysical research communications. 2000;273(1):239-45.

60. Lahortiga I, Vazquez I, Agirre X, Larrayoz MJ, Vizmanos JL, Gozzetti A, et al. Molecular heterogeneity in AML/MDS patients with 3q21q26 rearrangements. Genes, chromosomes & cancer. 2004;40(3):179-89.
61. Wlodarski MW, Hirabayashi S, Pastor V, Stary J, Hasle H, Masetti R, et al. Prevalence, clinical characteristics and prognosis of GATA2-related myelodysplastic syndromes (MDS) in children and adolescents. Blood. 2015.
62. Zhang SJ, Ma LY, Huang QH, Li G, Gu BW, Gao XD, et al. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(6):2076-81.
63. Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. Blood. 2011;118(10):2656-8.
64. Bigley V, Collin M. Dendritic cell, monocyte, B and NK lymphoid deficiency defines the lost lineages of a new GATA-2 dependent myelodysplastic syndrome. Haematologica. 2011;96(8):1081-3.
65. Pasquet M, Bellanne-Chantelot C, Tavitian S, Prade N, Beaupain B, Larochelle O, 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):822-9.
66. Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, Hickstein DD, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood. 2013.
67. Young NS. Acquired aplastic anemia. Annals of internal medicine. 2002;136(7):534-46.
68. Fujimaki S, Harigae H, Sugawara T, Takasawa N, Sasaki T, Kaku M. Decreased expression of transcription factor GATA-2 in haematopoietic stem cells in patients with aplastic anaemia. British journal of haematology. 2001;113(1):52-7.
69. Xu Y, Takahashi Y, Wang Y, Hama A, Nishio N, Muramatsu H, et al. Downregulation of GATA-2 and overexpression of adipogenic gene-PPARgamma in mesenchymal stem cells from patients with aplastic anemia. Experimental hematology. 2009;37(12):1393-9.
70. Townsley DM, Hsu AP, Dumitriu B, Holland SM, Young NS. Regulatory mutations in GATA2 associated with aplastic anemia. ASH Abstract. 2012.
71. Pang WW, Pluvinage JV, Price EA, Sridhar K, Arber DA, Greenberg PL, et al. Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(8):3011-6.
72. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nature genetics. 2011;43(4):309-15.
73. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
74. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
75. Koschmieder S, Halmos B, Levantini E, Tenen DG. Dysregulation of the C/EBPalpha differentiation pathway in human cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(4):619-28.

76. Nanri T, Uike N, Kawakita T, Iwanaga E, Mitsuya H, Asou N. 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):237-41.
77. Roumier C, Fenaux P, Lafage M, Imbert M, Eclache V, Preudhomme C. New mechanisms of AML1 gene alteration in hematological malignancies. Leukemia. 2003;17(1):9-16.
78. Christiansen DH, Andersen MK, Pedersen-Bjergaard J. Mutations with loss of heterozygosity of p53 are common in therapy-related myelodysplasia and acute myeloid leukemia after exposure to alkylating agents and significantly associated with deletion or loss of 5q, a complex karyotype, and a poor prognosis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2001;19(5):1405-13.
79. Guillamot M, Cimmino L, Aifantis I. The Impact of DNA Methylation in Hematopoietic Malignancies. Trends in cancer. 2016;2(2):70-83.
80. Greenblatt SM, Nimer SD. Chromatin modifiers and the promise of epigenetic therapy in acute leukemia. Leukemia. 2014;28(7):1396-406.
81. Hahn CN, Venugopal P, Scott HS, Hiwase DK. Splice factor mutations and alternative splicing as drivers of hematopoietic malignancy. Immunological reviews. 2015;263(1):257-78.
82. Boissel N, Leroy H, Brethon B, Philippe N, de Botton S, Auvrignon A, et al. Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). Leukemia. 2006;20(6):965-70.
83. Abu-Duhier FM, Goodeve AC, Wilson GA, Gari MA, Peake IR, Rees DC, et al. FLT3 internal tandem duplication mutations in adult acute myeloid leukaemia define a high-risk group. British journal of haematology. 2000;111(1):190-5.
84. Thol F, Bollin R, Gehlhaar M, Walter C, Dugas M, Suchanek KJ, et al. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. Blood. 2014;123(6):914-20.
85. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood. 2012;119(14):3203-10.
86. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-7.
87. De Keersmaecker K, Sulima SO, Dinman JD. Ribosomopathies and the paradox of cellular hypo- to hyperproliferation. Blood. 2015;125(9):1377-82.
88. Doherty L, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Clinton C, et al. Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. American journal of human genetics. 2010;86(2):222-8.
89. Narla A, Ebert BL. Ribosomopathies: human disorders of ribosome dysfunction. Blood. 2010;115(16):3196-205.
90. Faber K, Bullinger L, Ragu C, Garding A, Mertens D, Miller C, et al. CDX2-driven leukemogenesis involves KLF4 repression and deregulated PPARgamma signaling. The Journal of clinical investigation. 2013;123(1):299-314.
91. Yu Y, Yip KH, Tam IY, Sam SW, Ng CW, Zhang W, et al. Differential effects of the Toll-like receptor 2 agonists, PGN and Pam3CSK4 on anti-IgE induced human mast cell activation. PloS one. 2014;9(11):e112989.

92. Haugas M, Lillevali K, Hakanen J, Salminen M. Gata2 is required for the development of inner ear semicircular ducts and the surrounding perilymphatic space. Developmental dynamics : an official publication of the American Association of Anatomists. 2010;239(9):2452-69.
93. Dahlin JS, Hallgren J. Mast cell progenitors: origin, development and migration to tissues. Molecular immunology. 2015;63(1):9-17.
94. Iwasaki H, Mizuno S, Arinobu Y, Ozawa H, Mori Y, Shigematsu H, et al. The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. Genes & development. 2006;20(21):3010-21.
95. Urb M, Sheppard DC. The role of mast cells in the defence against pathogens. PLoS pathogens. 2012;8(4):e1002619.
96. Bergot AS, Kassianos A, Frazer IH, Mittal D. New Approaches to Immunotherapy for HPV Associated Cancers. Cancers. 2011;3(3):3461-95.
97. Yip KH, Kolesnikoff N, Yu C, Hauschild N, Taing H, Biggs L, et al. Mechanisms of vitamin D(3) metabolite repression of IgE-dependent mast cell activation. The Journal of allergy and clinical immunology. 2014;133(5):1356-64, 64 e1-14.
98. Raponi M, Lancet JE, Fan H, Dossey L, Lee G, Gojo I, et al. A 2-gene classifier for predicting response to the farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia. Blood. 2008;111(5):2589-96.
99. Tagliafico E, Tenedini E, Manfredini R, Grande A, Ferrari F, Roncaglia E, et al. Identification of a molecular signature predictive of sensitivity to differentiation induction in acute myeloid leukemia. Leukemia. 2006;20(10):1751-8.
100. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453-74.
101. Rollig C, Bornhauser M, Thiede C, Taube F, Kramer M, Mohr B, et al. Long-term prognosis of acute myeloid leukemia according to the new genetic risk classification of the European LeukemiaNet recommendations: evaluation of the proposed reporting system. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(20):2758-65.
102. Mrozek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP, et al. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(36):4515-23.
103. Li Z, Herold T, He C, Valk PJ, Chen P, Jurinovic V, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31(9):1172-81.
104. Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K, et al. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia. 2014;28(8):1586-95.
105. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric

averaging of multiple internal control genes. Genome biology. 2002;3(7):RESEARCH0034.
106. Federici L, Falini B. Nucleophosmin mutations in acute myeloid leukemia: a tale of protein unfolding and mislocalization. Protein science : a publication of the Protein Society. 2013;22(5):545-56.
107. Falini B, Bolli N, Liso A, Martelli MP, Mannucci R, Pileri S, et al. Altered nucleophosmin transport in acute myeloid leukaemia with mutated NPM1: molecular basis and clinical implications. Leukemia. 2009;23(10):1731-43.
108. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer cell. 2010;17(1):13-27.
109. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, et al. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. Blood. 2005;106(12):3747-54.
110. Mann RS, Lelli KM, Joshi R. Hox specificity unique roles for cofactors and collaborators. Current topics in developmental biology. 2009;88:63-101.
111. Agrawal-Singh S, Isken F, Agelopoulos K, Klein HU, Thoennissen NH, Koehler G, et al. Genome-wide analysis of histone H3 acetylation patterns in AML identifies PRDX2 as an epigenetically silenced tumor suppressor gene. Blood. 2012;119(10):2346-57.
112. Schnittger S, Bacher U, Haferlach C, Alpermann T, Dicker F, Sundermann J, et al. Characterization of NPM1-mutated AML with a history of myelodysplastic syndromes or myeloproliferative neoplasms. Leukemia. 2011;25(4):615-21.
113. Liu Y, He P, Liu F, Shi L, Zhu H, Zhao J, et al. Prognostic significance of NPM1 mutations in acute myeloid leukemia: A meta-analysis. Molecular and clinical oncology. 2014;2(2):275-81.
114. Schneider F, Hoster E, Unterhalt M, Schneider S, Dufour A, Benthaus T, et al. NPM1 but not FLT3-ITD mutations predict early blast cell clearance and CR rate in patients with normal karyotype AML (NK-AML) or high-risk myelodysplastic syndrome (MDS). Blood. 2009;113(21):5250-3.
115. Mullighan CG, Kennedy A, Zhou X, Radtke I, Phillips LA, Shurtleff SA, et al. Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. Leukemia. 2007;21(9):2000-9.
116. Jost E, Lin Q, Weidner CI, Wilop S, Hoffmann M, Walenda T, et al. Epimutations mimic genomic mutations of DNMT3A in acute myeloid leukemia. Leukemia. 2014;28(6):1227-34.
117. Khan SN, Jankowska AM, Mahfouz R, Dunbar AJ, Sugimoto Y, Hosono N, et al. Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. Leukemia. 2013;27(6):1301-9.
118. Roy DM, Walsh LA, Chan TA. Driver mutations of cancer epigenomes. Protein & cell. 2014;5(4):265-96.
119. Taoudi S, Bee T, Hilton A, Knezevic K, Scott J, Willson TA, et al. ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. Genes & development. 2011;25(3):251-62.

120. Diffner E, Beck D, Gudgin E, Thoms JA, Knezevic K, Pridans C, et al. Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. Blood. 2013;121(12):2289-300.
121. Loughran SJ, Kruse EA, Hacking DF, de Graaf CA, Hyland CD, Willson TA, et al. The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nature immunology. 2008;9(7):810-9.
122. Huang Y, Thoms JA, Tursky ML, Knezevic K, Beck D, Chandrakanthan V, et al. MAPK/ERK2 phosphorylates ERG at serine 283 in leukemic cells and promotes stem cell signatures and cell proliferation. Leukemia. 2016;30(7):1552-61.
123. Cutts BA, Sjogren AK, Andersson KM, Wahlstrom AM, Karlsson C, Swolin B, et al. Nf1 deficiency cooperates with oncogenic K-RAS to induce acute myeloid leukemia in mice. Blood. 2009;114(17):3629-32.
124. Langendijk JA, Slotman BJ, van der Waal I, Doornaert P, Berkof J, Leemans CR. Risk-group definition by recursive partitioning analysis of patients with squamous cell head and neck carcinoma treated with surgery and postoperative radiotherapy. Cancer. 2005;104(7):1408-17.
125. Ko GM, Reddy AS, Kumar S, Bailey BA, Garg R. Computational analysis of HIV-1 protease protein binding pockets. J Chem Inf Model. 2010;50(10):1759-71.
126. Omurlua IK, Turea M, Tokatli F. The comparisons of random survival forests and Cox regression analysis with simulation and an application related to breast cancer. Expert Systems with Applications. 2009;36:8582-8.
127. Pigazzi M, Masetti R, Martinolli F, Manara E, Beghin A, Rondelli R, et al. Presence of high-ERG expression is an independent unfavorable prognostic marker in MLL-rearranged childhood myeloid leukemia. Blood. 2012;119(4):1086-7; author reply 7-8.
128. Schwind S, Marcucci G, Maharry K, Radmacher MD, Mrozek K, Holland KB, et al. BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. Blood. 2010;116(25):5660-9.
129. Kornblau SM, Qiu YH, Zhang N, Singh N, Faderl S, Ferrajoli A, et al. Abnormal expression of FLI1 protein is an adverse prognostic factor in acute myeloid leukemia. Blood. 2011;118(20):5604-12.
130. Ohanian M, Rozovski U, Kantarjian H, Loghavi S, Huh Y, Abruzzo L, et al. MYC Expression Is Prognostic in Therapy Related Acute Myeloid Leukemia (AML) and AML with Myelodysplastic Syndrome (MDS)-Related Changes. Blood. 2014;124(1).
131. Li L, Wei XH, Pan YP, Li HC, Yang H, He QH, et al. LAPTM4B: a novel cancer-associated gene motivates multidrug resistance through efflux and activating PI3K/AKT signaling. Oncogene. 2010;29(43):5785-95.
132. Li Y, Iglehart JD, Richardson AL, Wang ZC. The amplified cancer gene LAPTM4B promotes tumor growth and tolerance to stress through the induction of autophagy. Autophagy. 2012;8(2):273-4.
133. Zhang H, Wei Q, Liu R, Qi S, Liang P, Qi C, et al. Overexpression of LAPTM4B-35: a novel marker of poor prognosis of prostate cancer. PloS one. 2014;9(3):e91069.

134. Tang H, Tian H, Yue W, Li L, Li S, Gao C, et al. Overexpression of LAPTM4B is correlated with tumor angiogenesis and poor prognosis in non-small cell lung cancer. Medical oncology. 2014;31(6):974.
135. Morad SA, Cabot MC. Ceramide-orchestrated signalling in cancer cells. Nature reviews Cancer. 2013;13(1):51-65.
136. Zhang H, Mi JQ, Fang H, Wang Z, Wang C, Wu L, et al. Preferential eradication of acute myelogenous leukemia stem cells by fenretinide. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(14):5606-11.
137. Maas AH, van der Schouw YT, Beijerinck D, Deurenberg JJ, Mali WP, van der Graaf Y. Arterial calcium on mammograms is not associated with inflammatory markers for heart disease risk. Heart. 2006;92(4):541-2.
138. Pesakhov S, Nachliely M, Barvish Z, Aqaqe N, Schwartzman B, Voronov E, et al. Cancer-selective cytotoxic Ca2+ overload in acute myeloid leukemia cells and attenuation of disease progression in mice by synergistically acting polyphenols curcumin and carnosic acid. Oncotarget. 2016.
139. Ma S, Jiang B, Deng W, Gu ZK, Wu FZ, Li T, et al. D-2-hydroxyglutarate is essential for maintaining oncogenic property of mutant IDH-containing cancer cells but dispensable for cell growth. Oncotarget. 2015;6(11):8606-20.
140. Hebestreit K, Grottrup S, Emden D, Veerkamp J, Ruckert C, Klein HU, et al. Leukemia gene atlas--a public platform for integrative exploration of genome-wide molecular data. PloS one. 2012;7(6):e39148.
141. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. The New England journal of medicine. 2004;350(16):1617-28.
142. Arozarena I, Calvo F, Crespo P. Ras, an actor on many stages: posttranslational modifications, localization, and site-specified events. Genes & cancer. 2011;2(3):182-94.
143. Cox AD, Der CJ. Ras history: The saga continues. Small GTPases. 2010;1(1):2-27.
144. Parikh C, Subrahmanyam R, Ren R. Oncogenic NRAS rapidly and efficiently induces CMML- and AML-like diseases in mice. Blood. 2006;108(7):2349-57.
145. Zhang J, Liu Y, Beard C, Tuveson DA, Jaenisch R, Jacks TE, et al. Expression of oncogenic K-ras from its endogenous promoter leads to a partial block of erythroid differentiation and hyperactivation of cytokine-dependent signaling pathways. Blood. 2007;109(12):5238-41.
146. Steckel M, Molina-Arcas M, Weigelt B, Marani M, Warne PH, Kuznetsov H, et al. Determination of synthetic lethal interactions in KRAS oncogene-dependent cancer cells reveals novel therapeutic targeting strategies. Cell research. 2012;22(8):1227-45.
147. Kumar MS, Hancock DC, Molina-Arcas M, Steckel M, East P, Diefenbacher M, et al. The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. Cell. 2012;149(3):642-55.
148. McCormack MP, Shields BJ, Jackson JT, Nasa C, Shi W, Slater NJ, et al. Requirement for Lyl1 in a model of Lmo2-driven early T-cell precursor ALL. Blood. 2013;122(12):2093-103.

149. Lassen LB, Ballarin-Gonzalez B, Schmitz A, Fuchtbauer A, Pedersen FS, Fuchtbauer EM. Nras overexpression results in granulocytosis, T-cell expansion and early lethality in mice. PloS one. 2012;7(8):e42216.
150. Li Q, Haigis KM, McDaniel A, Harding-Theobald E, Kogan SC, Akagi K, et al. Hematopoiesis and leukemogenesis in mice expressing oncogenic NrasG12D from the endogenous locus. Blood. 2011;117(6):2022-32.
151. Gao J, Chen YH, Peterson LC. GATA family transcriptional factors: emerging suspects in hematologic disorders. Experimental hematology & oncology. 2015;4:28.
152. Socolovsky M, Lodish HF, Daley GQ. Control of hematopoietic differentiation: lack of specificity in signaling by cytokine receptors. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(12):6573-5.
153. Parikh C, Subrahmanyam R, Ren R. Oncogenic NRAS, KRAS, and HRAS exhibit different leukemogenic potentials in mice. Cancer research. 2007;67(15):7139-46.
154. Wang J, Liu Y, Tan LX, Lo JC, Du J, Ryu MJ, et al. Distinct requirements of hematopoietic stem cell activity and Nras G12D signaling in different cell types during leukemogenesis. Cell cycle. 2011;10(17):2836-9.
155. Neubauer A, Maharry K, Mrozek K, Thiede C, Marcucci G, Paschka P, et al. Patients with acute myeloid leukemia and RAS mutations benefit most from postremission high-dose cytarabine: a Cancer and Leukemia Group B study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(28):4603-9.
156. Celton M, Forest A, Gosse G, Lemieux S, Hebert J, Sauvageau G, et al. Epigenetic regulation of GATA2 and its impact on normal karyotype acute myeloid leukemia. Leukemia. 2014;28(8):1617-26.
157. West RR, Hsu AP, Holland SM, Cuellar-Rodriguez J, Hickstein DD. Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. Haematologica. 2014;99(2):276-81.
158. Holme H, Hossain U, Kirwan M, Walne A, Vulliamy T, Dokal I. Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. British journal of haematology. 2012;158(2):242-8.
159. Mendola A, Schlogel MJ, Ghalamkarpour A, Irrthum A, Nguyen HL, Fastre E, et al. Mutations in the VEGFR3 signaling pathway explain 36% of familial lymphedema. Mol Syndromol. 2013;4(6):257-66.
160. Spinner MA, Sanchez LA, Hsu AP, Shaw PA, Zerbe CS, Calvo KR, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. Blood. 2014;123(6):809-21.
161. Ganapathi KA, Townsley DM, Hsu AP, Arthur DC, Zerbe CS, Cuellar-Rodriguez J, et al. GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia. Blood. 2015;125(1):56-70.
162. Vinh DC, Patel SY, Uzel G, Anderson VL, Freeman AF, Olivier KN, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. Blood. 2010;115(8):1519-29.
163. Dickinson RE, Milne P, Jardine L, Zandi S, Swierczek SI, McGovern N, et al. The evolution of cellular deficiency in GATA2 mutation. Blood. 2014;123(6):863-74.

164. Mace EM, Hsu AP, Monaco-Shawver L, Makedonas G, Rosen JB, Dropulic L, et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset. Blood. 2013;121(14):2669-77.
165. Green CL, Tawana K, Hills RK, Bodor C, Fitzgibbon J, Inglott S, et al. GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. British journal of haematology. 2013;161(5):701-5.
166. Shiba N, Funato M, Ohki K, Park MJ, Mizushima Y, Adachi S, et al. Mutations of the GATA2 and CEBPA genes in paediatric acute myeloid leukaemia. British journal of haematology. 2014;164(1):142-5.
167. Bigley V, Haniffa M, Doulatov S, Wang XN, Dickinson R, McGovern N, et al. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. The Journal of experimental medicine. 2011;208(2):227-34.
168. Wang X, Muramatsu H, Okuno Y, Sakaguchi H, Yoshida K, Kawashima N, et al. GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. Haematologica. 2015;100(10):e398-401.
169. Walter MJ, Shen D, Shao J, Ding L, White BS, Kandoth C, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia. 2013;27(6):1275-82.
170. Theis F, Corbacioglu A, Gaidzik VI, Paschka P, Weber D, Bullinger L, et al. Clinical impact of GATA2 mutations in acute myeloid leukemia patients harboring CEBPA mutations: a study of the AML study group. Leukemia. 2016;30(11):2248-50.
171. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Gorlich D, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. Blood. 2016;128(5):686-98.
172. Hou HA, Lin YC, Kuo YY, Chou WC, Lin CC, Liu CY, et al. GATA2 mutations in patients with acute myeloid leukemia-paired samples analyses show that the mutation is unstable during disease evolution. Ann Hematol. 2015;94(2):211-21.
173. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122(22):3616-27; quiz 99.
174. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481(7380):157-63.
175. Fasan A, Eder C, Haferlach C, Grossmann V, Kohlmann A, Dicker F, 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):482-5.
176. Grossmann V, Haferlach C, Nadarajah N, Fasan A, Weissmann S, Roller A, et al. CEBPA double-mutated acute myeloid leukaemia harbours concomitant molecular mutations in 76.8% of cases with TET2 and GATA2 alterations impacting prognosis. British journal of haematology. 2013;161(5):649-58.
177. Niimi K, Kiyoi H, Ishikawa Y, Hayakawa F, Kurahashi S, Kihara R, et al. GATA2 zinc finger 2 mutation found in acute myeloid leukemia impairs myeloid differentiation. Leuk Res Rep. 2013;2(1):21-5.
178. Dolnik A, Engelmann JC, Scharfenberger-Schmeer M, Mauch J, Kelkenberg-Schade S, Haldemann B, et al. Commonly altered genomic regions in acute myeloid

leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. Blood. 2012;120(18):e83-92.
179. Dufour A, Konstandin N, Ksienzyk B, Zellmeier E, Benthaus T, Yaghmaie M, et al. High Frequency of GATA2 Mutations in Cytogenetically Normal Acute Myeloid Leukemia with Biallelic CEBPA Mutations Identified by Exome Sequencing. Blood. 2011;118(21):72-.
180. Greif PA, Dufour A, Konstandin NP, Ksienzyk B, Zellmeier E, Tizazu B, 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):395-403.
181. Fujiwara T, Fukuhara N, Funayama R, Nariai N, Kamata M, Nagashima T, et al. Identification of acquired mutations by whole-genome sequencing in GATA-2 deficiency evolving into myelodysplasia and acute leukemia. Ann Hematol. 2014;93(9):1515-22.
182. . !!! INVALID CITATION !!! [].
183. Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, Hickstein DD, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood. 2013;121(19):3830-7, S1-7.
184. Kaur J, Catovsky D, Valdimarsson H, Jensson O, Spiers AS. Familial acute myeloid leukaemia with acquired Pelger-Huet anomaly and aneuploidy of C group. Br Med J. 1972;4(5836):327-31.
185. Bodor C, Renneville A, Smith M, Charazac A, Iqbal S, Etancelin P, 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):890-4.
186. Svobodova T, Mejstrikova E, Salzer U, Sukova M, Hubacek P, Matej R, et al. Diffuse parenchymal lung disease as first clinical manifestation of GATA-2 deficiency in childhood. BMC Pulm Med. 2015;15:8.
187. Alter BP, Giri N, Calvo KR, Maric I, Arthur DC, Hickstein DD, et al. Clinically Silent Carriers in Families with Myelodysplastic Syndrome Due to GATA2 Mutations. Blood. 2012;120(21):1264-.
188. Cuellar-Rodriguez J, Gea-Banacloche J, Freeman AF, Hsu AP, Zerbe CS, Calvo KR, et al. Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency. Blood. 2011;118(13):3715-20.
189. Ishida H, Imai K, Honma K, Tamura S, Imamura T, Ito M, 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):1273-6.
190. Camargo JF, Lobo SA, Hsu AP, Zerbe CS, Wormser GP, Holland SM. MonoMAC syndrome in a patient with a GATA2 mutation: case report and review of the literature. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2013;57(5):697-9.
191. Lubking A, Vosberg S, Konstandin NP, Dufour A, Graf A, Krebs S, et al. Young woman with mild bone marrow dysplasia, GATA2 and ASXL1 mutation treated with allogeneic hematopoietic stem cell transplantation. Leuk Res Rep. 2015;4(2):72-5.
192. Johnson KD, Hsu AP, Ryu MJ, Wang J, Gao X, Boyer ME, et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. The Journal of clinical investigation. 2012;122(10):3692-704.

193. Mutsaers PG, van de Loosdrecht AA, Tawana K, Bodor C, Fitzgibbon J, Menko FH. Highly variable clinical manifestations in a large family with a novel GATA2 mutation. Leukemia. 2013;27(11):2247-8.
194. Stieglitz E, Liu YL, Emanuel PD, Castleberry RP, Cooper TM, Shannon KM, et al. Mutations in GATA2 are rare in juvenile myelomonocytic leukemia. Blood. 2014;123(9):1426-7.
195. He B, Lanz RB, Fiskus W, Geng C, Yi P, Hartig SM, et al. GATA2 facilitates steroid receptor coactivator recruitment to the androgen receptor complex. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(51):18261-6.
196. Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. Journal of hematology & oncology. 2012;5:12.
197. Weber S, Alpermann T, Dicker F, Jeromin S, Nadarajah N, Eder C, et al. BAALC expression: a suitable marker for prognostic risk stratification and detection of residual disease in cytogenetically normal acute myeloid leukemia. Blood cancer journal. 2014;4:e173.
198. Abdel-Wahab O, Dey A. The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics. Leukemia. 2013;27(1):10-5.
199. Sawa M, Yamamoto K, Yokozawa T, Kiyoi H, Hishida A, Kajiguchi T, et al. BMI-1 is highly expressed in M0-subtype acute myeloid leukemia. International journal of hematology. 2005;82(1):42-7.
200. Dang H, Chen Y, Kamel-Reid S, Brandwein J, Chang H. CD34 expression predicts an adverse outcome in patients with NPM1-positive acute myeloid leukemia. Human pathology. 2013;44(10):2038-46.
201. Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton S, et al. Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). Blood. 2002;100(8):2717-23.
202. Kornblau SM, McCue D, Singh N, Chen W, Estrov Z, Coombes KR. Recurrent expression signatures of cytokines and chemokines are present and are independently prognostic in acute myelogenous leukemia and myelodysplasia. Blood. 2010;116(20):4251-61.
203. Linnemann AK, O'Geen H, Keles S, Farnham PJ, Bresnick EH. Genetic framework for GATA factor function in vascular biology. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(33):13641-6.
204. Dirksen U, Hattenhorst U, Schneider P, Schroten H, Gobel U, Bocking A, et al. Defective expression of granulocyte-macrophage colony-stimulating factor/interleukin-3/interleukin-5 receptor common beta chain in children with acute myeloid leukemia associated with respiratory failure. Blood. 1998;92(4):1097-103.
205. Ouboussad L, Kreuz S, Lefevre PF. CTCF depletion alters chromatin structure and transcription of myeloid-specific factors. Journal of molecular cell biology. 2013;5(5):308-22.
206. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. The New England journal of medicine. 2010;363(25):2424-33.

207. Rucker FG, Sander S, Dohner K, Dohner H, Pollack JR, Bullinger L. Molecular profiling reveals myeloid leukemia cell lines to be faithful model systems characterized by distinct genomic aberrations. Leukemia. 2006;20(6):994-1001.
208. Lund K, Adams PD, Copland M. EZH2 in normal and malignant hematopoiesis. Leukemia. 2014;28(1):44-9.
209. Levis M, Small D. FLT3: ITDoes matter in leukemia. Leukemia. 2003;17(9):1738-52.
210. Khanim FL, Bradbury CA, Arrazi J, Hayden RE, Rye A, Basu S, et al. Elevated FOSB-expression; a potential marker of valproate sensitivity in AML. British journal of haematology. 2009;144(3):332-41.
211. Shimamoto T, Ohyashiki JH, Ohyashiki K, Kawakubo K, Kimura N, Nakazawa S, et al. GATA-1, GATA-2, and stem cell leukemia gene expression in acute myeloid leukemia. Leukemia. 1994;8(7):1176-80.
212. Magalhaes IQ, Splendore A, Emerenciano M, Figueiredo A, Ferrari I, Pombo-de-Oliveira MS. GATA1 mutations in acute leukemia in children with Down syndrome. Cancer genetics and cytogenetics. 2006;166(2):112-6.
213. Anguita E, Gupta R, Olariu V, Valk PJ, Peterson C, Delwel R, et al. A somatic mutation of GFI1B identified in leukemia alters cell fate via a SPI1 (PU.1) centered genetic regulatory network. Developmental biology. 2016;411(2):277-86.
214. van der Meer LT, Jansen JH, van der Reijden BA. Gfi1 and Gfi1b: key regulators of hematopoiesis. Leukemia. 2010;24(11):1834-43.
215. Thorsteinsdottir U, Mamo A, Kroon E, Jerome L, Bijl J, Lawrence HJ, et al. Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. Blood. 2002;99(1):121-9.
216. Lu Y, Wang Q, Mu QT, Chen ZM, Lou JY, Ni WM, et al. [Expression and clinical significance of ID1 gene in acute myeloid leukemia]. Zhonghua xue ye xue za zhi = Zhonghua xueyexue zazhi. 2012;33(4):278-81.
217. Dang L, Jin S, Su SM. IDH mutations in glioma and acute myeloid leukemia. Trends in molecular medicine. 2010;16(9):387-97.
218. de Rooij JD, Beuling E, van den Heuvel-Eibrink MM, Obulkasim A, Baruchel A, Trka J, et al. Recurrent deletions of IKZF1 in pediatric acute myeloid leukemia. Haematologica. 2015;100(9):1151-9.
219. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nature genetics. 2013;45(3):242-52.
220. Stirewalt DL, Meshinchi S, Kopecky KJ, Fan W, Pogosova-Agadjanyan EL, Engel JH, et al. Identification of genes with abnormal expression changes in acute myeloid leukemia. Genes, chromosomes & cancer. 2008;47(1):8-20.
221. Taskesen E, Havermans M, van Lom K, Sanders MA, van Norden Y, Bindels E, et al. Two splice-factor mutant leukemia subgroups uncovered at the boundaries of MDS and AML using combined gene expression and DNA-methylation profiling. Blood. 2014;123(21):3327-35.
222. Siatecka M, Bieker JJ. The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood. 2011;118(8):2044-54.
223. Metzeler KH, Heilmeier B, Edmaier KE, Rawat VP, Dufour A, Dohner K, et al. High expression of lymphoid enhancer-binding factor-1 (LEF1) is a novel favorable

prognostic factor in cytogenetically normal acute myeloid leukemia. Blood. 2012;120(10):2118-26.
224. Patel JL, Pournazari P, Haggstrom SJ, Kosari F, Shabani-Rad MT, Natkunam Y, et al. LMO2 (LIM domain only 2) is expressed in a subset of acute myeloid leukaemia and correlates with normal karyotype. Histopathology. 2014;64(2):226-33.
225. Meng YS, Khoury H, Dick JE, Minden MD. Oncogenic potential of the transcription factor LYL1 in acute myeloblastic leukemia. Leukemia. 2005;19(11):1941-7.
226. Wang QF, Li YJ, Dong JF, Li B, Kaberlein JJ, Zhang L, et al. Regulation of MEIS1 by distal enhancer elements in acute leukemia. Leukemia. 2014;28(1):138-46.
227. Munoz L, Nomdedeu JF, Villamor N, Guardia R, Colomer D, Ribera JM, et al. Acute myeloid leukemia with MLL rearrangements: clinicobiological features, prognostic impact and value of flow cytometry in the detection of residual leukemic cells. Leukemia. 2003;17(1):76-82.
228. Heuser M, Beutel G, Krauter J, Dohner K, von Neuhoff N, Schlegelberger B, et al. High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. Blood. 2006;108(12):3898-905.
229. Lorenzo PI, Brendeford EM, Gilfillan S, Gavrilov AA, Leedsak M, Razin SV, et al. Identification of c-Myb Target Genes in K562 Cells Reveals a Role for c-Myb as a Master Regulator. Genes & cancer. 2011;2(8):805-17.
230. Pattabiraman DR, McGirr C, Shakhbazov K, Barbier V, Krishnan K, Mukhopadhyay P, et al. Interaction of c-Myb with p300 is required for the induction of acute myeloid leukemia (AML) by human AML oncogenes. Blood. 2014;123(17):2682-90.
231. Salvatori B, Iosue I, Djodji Damas N, Mangiavacchi A, Chiaretti S, Messina M, et al. Critical Role of c-Myc in Acute Myeloid Leukemia Involving Direct Regulation of miR-26a and Histone Methyltransferase EZH2. Genes & cancer. 2011;2(5):585-92.
232. Kiani A, Rao A, Aramburu J. Manipulating immune responses with immunosuppressive agents that target NFAT. Immunity. 2000;12(4):359-72.
233. Robert-Moreno A, Espinosa L, de la Pompa JL, Bigas A. RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. Development. 2005;132(5):1117-26.
234. Wenzl K, Troppan K, Neumeister P, Deutsch AJ. The nuclear orphan receptor NR4A1 and NR4A3 as tumor suppressors in hematologic neoplasms. Current drug targets. 2015;16(1):38-46.
235. Tsuzuki S, Towatari M, Saito H, Enver T. 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. Molecular and cellular biology. 2000;20(17):6276-86.
236. Steinbach D, Hermann J, Viehmann S, Zintl F, Gruhn B. Clinical implications of PRAME gene expression in childhood acute myeloid leukemia. Cancer genetics and cytogenetics. 2002;133(2):118-23.
237. Guo L, Chen L, Bi S, Chai L, Wang Z, Cao C, et al. PTEN inhibits proliferation and functions of hypertrophic scar fibroblasts. Molecular and cellular biochemistry. 2012;361(1-2):161-8.
238. Nickels EM, Soodalter J, Churpek JE, Godley LA. Recognizing familial myeloid leukemia in adults. Therapeutic advances in hematology. 2013;4(4):254-69.

239. Score J, Hidalgo-Curtis C, Jones AV, Winkelmann N, Skinner A, Ward D, et al. Inactivation of polycomb repressive complex 2 components in myeloproliferative and myelodysplastic/myeloproliferative neoplasms. Blood. 2012;119(5):1208-13.
240. Weissmann S, Alpermann T, Grossmann V, Kowarsch A, Nadarajah N, Eder C, et al. Landscape of TET2 mutations in acute myeloid leukemia. Leukemia. 2012;26(5):934-42.
241. Hou HA, Chou WC, Kuo YY, Liu CY, Lin LI, Tseng MH, et al. TP53 mutations in de novo acute myeloid leukemia patients: longitudinal follow-ups show the mutation is stable during disease evolution. Blood cancer journal. 2015;5:e331.
242. Fathi AT, Abdel-Wahab O. Mutations in epigenetic modifiers in myeloid malignancies and the prospect of novel epigenetic-targeted therapy. Advances in hematology. 2012;2012:469592.
243. Aguayo A, Estey E, Kantarjian H, Mansouri T, Gidel C, Keating M, et al. Cellular vascular endothelial growth factor is a predictor of outcome in patients with acute myeloid leukemia. Blood. 1999;94(11):3717-21.
244. Tessema M, Langer F, Bock O, Seltsam A, Metzig K, Hasemeier B, et al. Down-regulation of the IGF-2/H19 locus during normal and malignant hematopoiesis is independent of the imprinting pattern. International journal of oncology. 2005;26(2):499-507.
245. Freson K, Thys C, Wittewrongel C, Vermylen J, Hoylaerts MF, Van Geet C. Molecular cloning and characterization of the GATA1 cofactor human FOG1 and assessment of its binding to GATA1 proteins carrying D218 substitutions. Human genetics. 2003;112(1):42-9.
246. Tevosian SG, Deconinck AE, Cantor AB, Rieff HI, Fujiwara Y, Corfas G, et al. FOG-2: A novel GATA-family cofactor related to multitype zinc-finger proteins Friend of GATA-1 and U-shaped. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(3):950-5.