How do Gata1 and cohesin gene mutations contribute to the development of myeloid leukaemia?

Catherine Garnett

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019

This thesis is dedicated to my family

"We shall not cease from exploration, and the end of all our exploring will be to arrive where we started and know the place for the first time."

T. S. Eliot

Acknowledgements

I would like to start by thanking my supervisors, Paresh Vyas and Irene Roberts, for their guidance, encouragement and support - not just over the course of my DPhil but from my first taste of Gata1 biology as a very junior doctor. I am grateful to Paresh for challenging me to develop independent thinking, to really understand the science underlying any experiment, and to always strive for the highest of standards. And to Irene, who has never failed to help me ask the right scientific questions, has been a constant source of thoughtful, insightful and practical advice and has inspired me through her passion for her work.

I would also like to thank all the members of their laboratories (past and present) who have helped me to learn techniques, provided helpful comments on my data and been there to listen when things went wrong. In particular Helen Doolittle, Jess Doondea, Batchimeg Usukhbayar, Hedia Chagroui, Marlen Metzner, Alison Kennedy, Maiken Kristiansen, Kelly Soady, David Cruz Hernandez, Anindita Roy, Sorcha O'Byrne, Nick Fordham, Natalina Elliot, I-Jun Lau, Zahra Aboukhalil, Nathalie Sakakini, Gaetan Juban and Ludovic Lhermitte.

I have benefited massively from mentorship, insightful discussions and practical guidance from Lynn Quek, Beth Psaila and James Davies - all of whom have enabled me to view my data in new and exciting ways and taken time out of their busy schedules to help me. I would like to thank Charlotte George, Katy Brown and Supat Thongjuea for their computational help and patience and Lars Hassen, Andy King, Damien Downes, Rob Beagrie, Dimitra Dimou, Deena Iskander and Catherine Porcher for discussion, protocols and advice.

I would like to acknowledge Doug Higgs, Claus Nerlov and Tom Milne for their advice and support in the capacity of my thesis committee and all the core facilities of the WIMM. It has been a privilege to train in such a stimulating and collaborative research environment.

This work was made possible by generous funding from the Wellcome Trust and an Anya Sturdy Fellowship for which I am very grateful.

Finally, I would like to thank my family. My parents, Amanda and Martin who have provided me with so much unconditional love, support and encouragement over the years and always believed that I would achieve whatever I set out to do. The Bousfield and Kerss families who have welcomed me into their lives with such generosity and been a constant source of emotional and practical support. And last but not least, my husband, Andrew, and son, Sammy, who every day inspire me, challenge me and make it all worthwhile. Andy and Sammy, thank you for everything. This thesis is dedicated to you.

Abstract

Acute myeloid leukaemia (AML) is the most common aggressive human leukaemia. It is frequently characterised by the stepwise acquisition of multiple somatic mutations in stem and progenitor cells. The mechanisms by which these mutations interact to perturb normal cellular processes and drive transformation remains a key question in cancer biology.

This thesis explores the co-operation between two mutations which commonly co-occur in an acute megakaryoblastic leukaemia unique to children with Down Syndrome: *Gata1s* and haploinsufficiency of the core cohesin component *Rad21*.

Myeloid leukaemia of Down Syndrome (ML-DS) provides an excellent model in which to study multistep leukaemogenesis. It is biologically simpler than adult AML, comprising just three genetic events occurring in temporarily separable stages: (i) constitutive trisomy 21; (ii) an acquired truncating mutation in the transcription factor *GATA1 (GATA1s)* in fetal life, resulting in a preleukaemic condition, Transient Abnormal Myelopoiesis (TAM); (iii) additional somatic mutation(s) in the first 4 years of post natal life, transforming preleukaemic TAM clone(s) to ML-DS. Previous studies indicate that just one additional mutation in a cohesin complex gene may be sufficient for transformation, suggesting an important co-operative role.

To address this, I first confirmed the co-existence of *Gata1s* and cohesin gene mutations in a large cohort of patient samples and characterised the frequency and predicted functional effect of these variants. I then explored the impact of these mutations, both alone and in combination, using a transgenic mouse model system. Combined Gata1s/cohesin mutant mice displayed a novel haematopoietic phenotype, characterised by expansion of the bone marrow megakaryocyte progenitor (MkP) compartment and abnormal differentiation and proliferation of the cells within it. Integrated molecular analysis of the MkP revealed altered gene expression profiles and chromatin accessibility, with dysregulation of a number of relevant transcription factor pathways.

Significantly, both phenotypic and molecular analyses revealed a much greater effect of *Gata1s* plus cohesin haploinsufficiency than of either mutation alone, supporting a role for mutational co-operativity. The true nature of this interaction seems likely to be complex, potentially involving dysregulation of multiple targets and regulatory mechanisms within the cell.

Contents

List of Figures

List of Abbreviations

3C . . . . . Chromosome conformation capture

AMKL . . . . Acute megakaryoblastic leukaemia

AML . . . . . Acute myeloid leukaemia

ATAC . . . . . Assay for Transposase-Accessible Chromatin

BM . . . . . Bone marrow

bp . . . . . Base pair

CdLS . . . . . Cornelia de Lange Syndrome

CFU . . . . . Colony forming unit

ChIP . . . . . Chromatin immunoprecipitation

CLL . . . . . Chronic lymphocytic leukaemia

CLP . . . . . Common lymphoid progenitor

CMP . . . . . Common myeloid progenitor

Comb . . . . . Combined Gata1s/Rad21 mutant

CUT&RUN . Cleavage Under Targets and Relase using Nuclease

DHPLC . . . . Direct High Performance Liquid Chromatography

DLBCL . . . . Diffuse large B-cell lymphoma

DNA . . . . . Deoxyribonucleic acid

DS . . . . . Down Syndrome

EDTA . . . . . Ethylenediaminetetraacetic acid

FCS . . . . . Fetal Calf Serum

FL . . . . . Full length

FMO . . . . . Fluorescence-minus-one

FOG-1 . . . . . Friend of GATA1

GATA1s . . . GATA1 short

GSEA . . . . . Gene Set Enrichment Analysis

GMP . . . . . Granulocyte macrophage progenitor

HSC . . . . . Haematopoietic stem cell

IPS . . . . . Induced pluripotent stem cells [kb] Kilo base

Lin . . . . . Lineage

LMPP . . . . . Lymphoid-primed multipotent progenitor

LSK . . . . . Lin-Sca-1+Kit+

LT-HSC . . . . Long term haematopoietic stem cell

MCV . . . . . mean cell volume

MEP . . . . . Megakaryocyte-erythroid progenitor

MkP . . . . . Megakaryocyte progenitor

ML-DS . . . . Myeloid Leukaemia of Down Syndrome

MNC . . . . . Mononuclear cell

MPP . . . . . Multipotent progenitor

NGS . . . . . Next generation sequencing

PBS . . . . . Phosphate Buffered Saline

PCA . . . . . Principal component analysis

PCR . . . . . Polymerase Chain Reaction

PreMegE . . . Pre-megakaryocyte erythroid progenitor

SEM . . . . . Standard Error of the Mean

SNP . . . . . Single nucleotide polymorphism

ST-HSC . . . . Short term haematopoietic stem cell

TAD . . . . . Topologically associating domain

TAM . . . . . Transient Abnormal Myelopoiesis

T21 . . . . . Trisomy 21

VAF . . . . . Variant allele frequency

vWF . . . . . von Willebrand Factor

WT . . . . . Wild type

1 Introduction

Contents

1.1 Cooperating mutations in myeloid leukaemia

Cancer is a genetically complicated disease, frequently characterised by the stepwise acquisition of multiple mutations in stem and progenitor cells. What the function of these mutations is, and how they co-operate to drive tumourigenesis remains a key area of research. Blood cancers, such as acute myeloid leukaemia (AML), are often genetically simpler that solid tumours (Figure 1.1) and therefore provide a good model in which to start to address these questions.

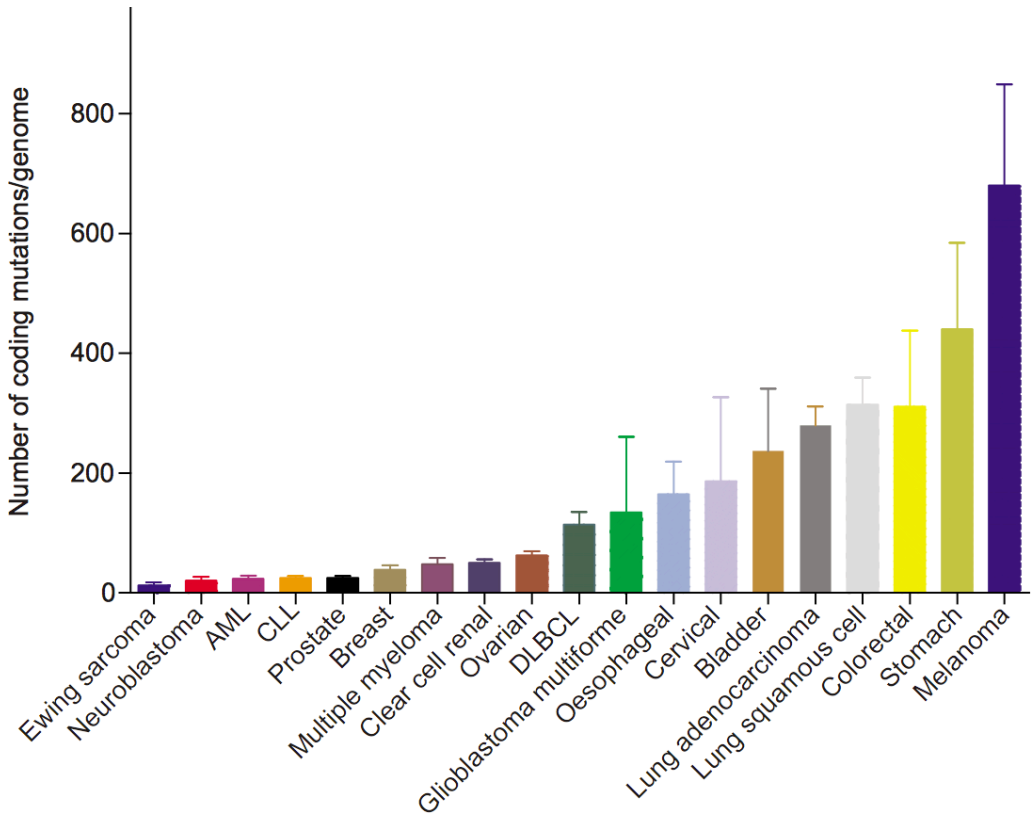


Figure 1.1: Acute myeloid leukaemia is genetically simpler than many solid tumours. Comparison of the mean number of coding mutations per genome in tumours of different tissues. Error bars show the standard error of the mean. AML, acute myeloid leukaemia; CLL, chronic lymphocytic leukaemia; DLBCL, diffuse large B-cell lymphoma. Figure and legend reproduced with permission from Grove and Vassiliou, 2014.

1.2 The genetic landscape of AML

Acute myeloid leukaemia (AML) is the most common aggressive human leukaemia, characterised by uncontrolled proliferation of abnormal haematopoietic progenitor cells and a block in myeloid differentiation, resulting in clinically significant cytopenias. Prognosis for patients is poor, with an overall 5-year survival of 10-40%. Death is often due to relapsed disease (Craddock et al. 2005, Grove et al. 2014).

The advent of sensitive Next Generation Sequencing (NGS) technologies has allowed detailed exploration of the mutational landscape of AML. In recent years there has been a shift towards classification and prognostication based on genetic data rather than morphology and cytogenetics. In 2013, Ley et al. reported whole genome/exome sequencing from 200 cases of adult de novo AML, revealing an

average of 13 coding mutations per patient, of which five were in recurrently mutated genes. Nearly all cases had at least one identifiable potential driver mutation and, in total, 259 genes were mutated in two or more samples. Categories of genes affected included tumour suppressors, genes associated with DNA methylation, *NPM1*, signalling molecules, chromatin modifiers, myeloid transcription factors, spliceosome-complex genes and cohesin complex genes. Interesting patterns of mutual exclusivity and mutational cooperation were noted (Ley 2013). More recently, Papaemmanuil et al. reported a comprehensive study of leukaemia genes in 1540 patients receiving intensive AML treatment. Over 5000 driver mutations were identified across 76 genomic regions, with 86% of patients having two or more drivers. Patients were then classified into 11 molecular subgroups based upon patterns of co-mutation, each with different characteristics and disease outcomes (Papaemmanuil et al. 2016). Given that multiple combinations of mutations may co-exist in different clones in an individual leukaemia patient, picking apart the cellular roles of these mutations is still far from straightforward.

1.3 Down Syndrome-associated myeloid disease

The unique natural history and relative genetic simplicity of a sub-type of AML found uniquely in children with Down Syndrome (constitutive Trisomy 21) make it an excellent model in which to study the cellular and molecular processes underlying mutational cooperation in multistep leukaemogenesis.

Despite not being cancer prone in general, children with Down Syndrome have a increased risk of developing acute myeloid or lymphoblastic leukaemia (~150 fold and 33 fold, respectively) (Hasle et al. 2000). Myeloid Leukaemia of Down Syndrome (ML-DS) is an acute megakaryocyte-erythroid leukaemia reported only during the first four years of life. It is genetically simpler even than adult AML, comprising just three, temporarily separable, stages of disease development (Figure 1.2):

1. Constitutive trisomy 21 (T21) leading to abnormal fetal haematopoiesis (Roy et al. 2012).

2. An acquired mutation in the key haematopoietic transcription factor *GATA1 (GATA1s)* in fetal life resulting in a preleukaemic condition, Transient Abnormal Myelopoiesis (TAM) (Ahmed et al. 2004, Alford et al. 2011, Roberts et al. 2013).

3. Additional somatic mutation(s) in post natal life, transforming preleukaemic TAM clone(s) to ML-DS (Nikolaev et al. 2013, Yoshida et al. 2013, Labuhn et al. 2019).

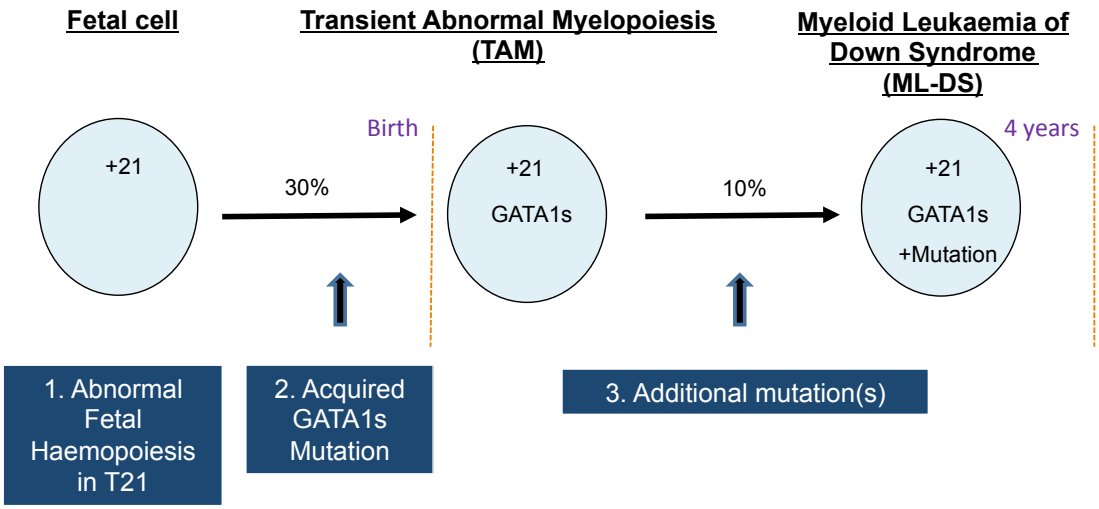


Figure 1.2: Myeloid Leukaemia of Down Syndrome (ML-DS): a model of multistep leukaemogenesis. Schematic of the genetic events leading to ML-DS. GATA1s = GATA1 short mutation, T21 = trisomy 21.

1.3.1 Transient Abnormal Myelopoiesis

Following the first report of somatic N-terminal truncating *GATA1* mutations in ML-DS (Wechsler et al. 2002), multiple reports confirmed these findings and showed that these mutations were also found in TAM. In neonates that developed TAM and later progressed to ML-DS, the TAM and ML-DS clones often had the same *GATA1* mutation showing that ML-DS directly evolves from TAM (Groet et al. 2003, Hitzler et al. 2003, Rainis et al. 2003, Ahmed et al. 2004, Alford et al. 2011). Around 25% of TAM and ML-DS cases had multiple *GATA1* mutant clones. Remarkably, a prospective population-based study identified *GATA1* mutations in nearly 1 in 3 Down Syndrome neonates at birth (Roberts et al. 2013).

TAM has a variable clinical presentation (Massey et al. 2006, Klusmann et al. 2008, Muramatsu et al. 2008). It is usually diagnosed at birth, however the diagnosis may be delayed and made at any stage within the first 8 weeks of life. The median age of diagnosis is 3-7 days. TAM is classically asymptomatic. However, it can present with jaundice, hepatomegaly, splenomegaly, skin rash and pleural/pericardial effusions. Laboratory tests may show leucocytosis, circulating blast cells and thrombocytopenia. Liver fibrosis, resulting from blast cell infiltration can lead to fulminant liver failure in a small number of cases. Recent UK guidelines recommend identifying those patients at high risk of early death (high white cell count, severe liver disease, presence of effusions and coagulopathies) and considering treatment with low dose cytarabine (Tunstall et al. 2018).

TAM has also been reported in fetal life, associated with hydrops fetalis or features mentioned above (Tamblyn et al. 2016). TAM blasts are typically megakaryoblastic and often co-express myeloid (CD33/CD13), megakaryocyte (CD42, CD61) and markers of immature cells (CD34, CD117) (Langebrake et al. 2005).

In keeping with previously published reports, data from the prospective multicentre Oxford Imperial Down Syndrome cohort study found that 10-15% of neonates with Down Syndrome were diagnosed with TAM in the context of *GATA1* mutation(s), circulating blasts of more than 10% and/or other clinical

features. An additional 10-15% of neonates with Down Syndrome were found to have acquired *GATA1* mutations but no clinical or haematological features of TAM (blasts <10%) and were classified as having silent TAM (Roberts et al. 2013).

All babies with TAM require regular clinical follow up and monitoring of their blood counts and films over the first 6 months. The majority of cases of both TAM and silent TAM resolve spontaneously without treatment and the majority of these children never develop ML-DS, (Klusmann et al. 2008, Gamis et al. 2011, Roberts et al. 2013).

1.3.2 Myeloid Leukaemia of Down Syndrome

Approximately 10-20% of neonates with TAM will go on to develop ML-DS (Klusmann et al. 2008, Gamis et al. 2011, Roberts et al. 2013). ML-DS is characterised molecularly by both the persistence of a *GATA1s* mutation (present in preceding TAM sample(s) but not necessarily the largest clone) and acquisition of additional co-operating mutation(s), most commonly in genes encoding components of the cohesin complex (*RAD21, STAG1/2, SMC1/3, CTCF*), epigenetic regulators (e.g. *KANSL1, EZH2,* and *SUZ12,*) or signalling molecules (e.g. *JAK* family, *MPL, KIT*) (Nikolaev et al. 2013, Yoshida et al. 2013, Labuhn et al. 2019).

The median age of diagnosis of ML-DS is 12-18 months and cases have not been reported beyond five years of age. ML-DS often presents with progressive pancytopenia and a relatively low circulating blast count. Bone marrow aspirate may give few cells due to marrow fibrosis. Blast cells (Figure 1.2) are morphologically and immunophenotypically similar to TAM blasts and carry *GATA1* mutations plus additional somatic variants. ML-DS patients usually respond well to chemotherapy, with most cases of treatment failure related to drug toxicity (Rao et al. 2006). Overall survival is better than for non-DS AML at around 80%. However treatment options for the 20% patients who relapse or fail to enter remission are limited (Rao et al. 2006, Gamis et al. 2011).

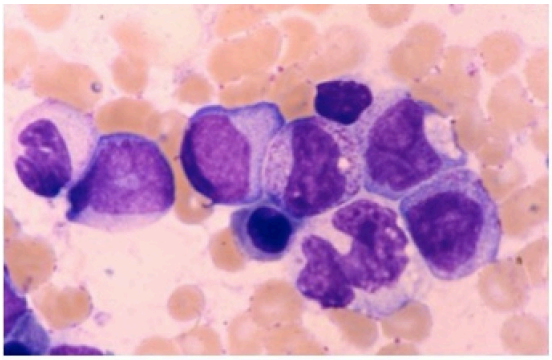


Figure 1.3: TAM/ML-DS peripheral blood smear. x100 magnification

1.4 GATA1: A key haematopoietic regulator

1.4.1 The GATA family of transcription factors

The founding member of the GATA family, GATA1, was first identified as a protein binding the 3' enhancer of the beta-globin gene (Wall et al. 1988) and cloned in 1989 (Tsai et al. 1989). Five further GATA family members (GATA 2-6) were subsequently identified (reviewed in (Bresnick et al. 2012)). These share two highly conserved zinc finger domains: a C-terminal zinc finger which is required for recognition and binding of the (A/T)GATA(A/G) consensus sequence (Yang and Evans, 1992) and an N-terminal zinc finger which is involved in stabilisation and recruitment of co-factors (Martin et al. 1990). All GATA family proteins also possess an N-terminal transactivation domain, however this shows considerable variation between factors (Martin et al. 1990, Molkentin 2000). Historically, the GATA factors have been divided into two subfamilies: GATA1, GATA2 and GATA3 which are expressed mainly in haematopoietic cells, but also endothelial cells and the nervous system (Nardelli et al. 1999), and GATA4, GATA5 and GATA6 which are expressed in mesoderm and endoderm-derived tissues including heart, liver, gut and gonad (Molkentin 2000).

1.4.2 GATA1 in normal haematopoiesis

It is well established that full length GATA1 is essential for normal erythropoiesis. GATA1 expression increases during erythroid lineage specification, but is down-

modulated in the late stages (Mouthon et al. 1993). Embryonic stem cells deficient in Gata1 are able to contribute to all tissues in chimeric mice, with the exception of mature erythroid cells (Pevny et al. 1991), while Gata1-null mice die at E10.5-E11.5 from fetal anaemia, with erythropoiesis arrested at the proerythroblast stage (Fujiwara et al. 1996). *In vitro* Gata1-deficient proerythroblasts demonstrated increased apoptosis (Weiss et al. 1995). "GATA1 low" mutant mice, expressing around 20% of wild type GATA1, were found to have a milder phenotype. In the majority of cases, this was still embryonic lethal, with death between E13.5-E14.5 due to ineffective erythropoiesis. However, a small number of Gata1-low mice survived to birth and those that did recover from their anaemia and were viable to adulthood (McDevitt et al. 1997). Conditional knockout of *Gata1* in adult mice resulted in a block in erythroid progenitor maturation. When treated with a haemolytic agent, these mice failed to mount an appropriate erythropoietic response to stress suggesting an essential role for Gata1 in both stress and steady-state erythropoiesis (Gutiérrez et al. 2007).

GATA1 is also critical for normal megakaryocyte development. Deletion of an upstream enhancer in the *Gata1* locus results in loss of megakaryocytic *Gata1* expression, a block in terminal megakaryopoiesis and marked thrombocytopenia (Shivdasani et al. 1997). Immature GATA1-deficient megakaryocyte progenitors accumulate in the spleen and bone marrow and show abnormal differentiation and proplatelet formation (Vyas et al. 1999). Important roles for GATA1 in eosinophil (Yu et al. 2002), mast cell (Migliaccio et al. 2003) and dendritic cell (Gutiérrez et al. 2007) differentiation have also been established.

To address the role of Gata1 in megakaryocyte/erythroid lineage specification, Mancini et al. acutely deleted Gata1 in a conditional mouse model. This did not affect the numbers of megakaryocyte/erythroid progenitors, but did completely block formation of early erythroid-commited progenitors, supporting a critical role for Gata1 in erythroid lineage commitment. Megakaryocyte progenitors were proportionately increased (Mancini et al. 2012). Significantly, the expression of transcription factors important to erythroid differentiation such as *Klf1, Gata2*

and Zfpm1 were not affected, suggesting a specific requirement for Gata1 in the megakaryocyte vs erythroid lineage decision.

The *GATA1* gene is located on the X chromosome and comprises 6 exons (Figure 1.4a). In normal human haematopoiesis, the GATA1 locus is alternatively spliced to produce two mRNA transcripts. One of the transcripts contains exons 1-6 and can potentially be translated into two protein isoforms, a full-length 413 amino-acid protein (GATA1fl) and a truncated 330 amino-acid isoform referred to as GATA1 short or GATA1s. GATA1s results from translation from an alternative ATG codon located in exon 3 and lacks the 83 amino acids comprising the N-terminal transactivation domain (Figure 1.4b) (Calligaris et al. 1995, Halsey et al. 2010). The second transcript lacks exon 2 but retains exon 1, and exons 3 to 6. This short transcript can only be translated into the GATA1s isoform. Both GATA1 isoforms contain two zinc finger domains, however only GATA1fl contains the classically defined transactivation domain. As yet, the role of GATA1s in normal haematopoiesis remains unknown.

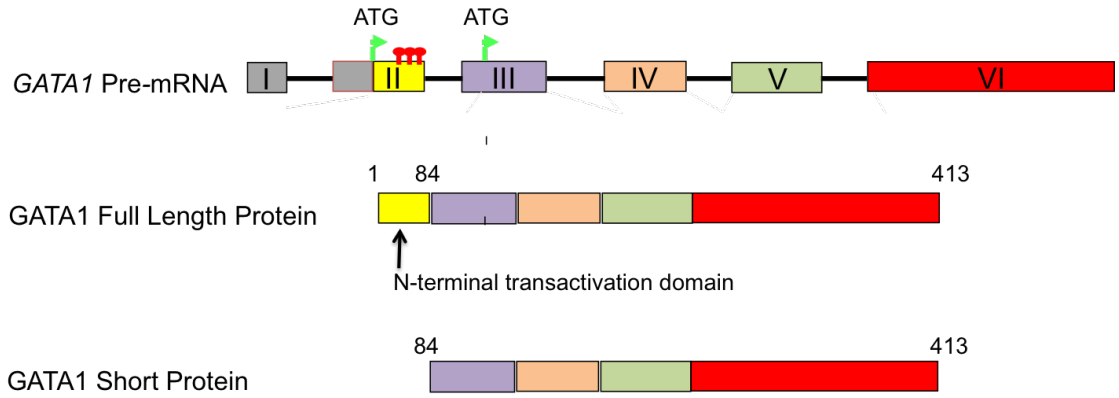


Figure 1.4: Gata1s: structure. Schematic showing mRNA transcript with alternative start codons (ATG) (top), full length protein (middle) and GATA1 short protein lacking the N-terminal transactivation domain. Blocks represent exons and black lines represent introns. Red markers show where the majority of *GATA1* mutations in TAM and ML-DS are found.

1.4.3 *GATA1* gene mutations in TAM and ML-DS

The somatic *GATA1* mutations reported in TAM and ML-DS occur at the 5' end of the gene, resulting in exclusive production of GATA1s in TAM and ML-DS

blasts (Wechsler et al. 2002, Rainis et al. 2003, Ahmed et al. 2004, Pine et al. 2007). These mutations prevent the synthesis of full length GATA-1 protein by impeding translation from the first ATG codon in exon 2. Instead, only the alternative ATG codon in exon 3 is used to translate the N-terminal–truncated GATA1s protein. Mutations in the *GATA1* locus that exclusively give rise to GATA1s are not found in non-Down Syndrome childhood leukaemias (Wechsler et al. 2002, Rainis et al. 2003), although they have been reported in cases of mosaic Down Syndrome (Reinhardt et al. 2010) and in association with acquired T21 (Rainis et al. 2003). The majority of these mutations are deletions, insertions and point mutations and are clustered in exon 2 or the beginning of exon 3. (Groet et al. 2003, Hitzler et al. 2003, Ahmed et al. 2004, Pine et al. 2007, Klusmann et al. 2008). These mutations are present in disease but not remission. The same *GATA1* mutations are found in TAM and ML-DS blasts, supporting a clonal relationship between the two conditions (Ahmed et al. 2004, Klusmann et al. 2008). In some patients, multiple *GATA1s* clones can be detected (Ahmed et al. 2004). However there does not appear to be a prognostic association between the number of *GATA1s* clones or type of *GATA1* mutations (Alford et al. 2011).

So how does exclusive production of GATA1s perturb haematopoiesis? There is growing evidence that the GATA1 N-terminus is important for terminal erythroid and megakaryocytic differentiation, particularly during fetal haematopoiesis. Early transfection experiments suggested that the GATA1 N-terminus acts as a transactivation domain, binding co-factors to augment GATA1 activity (Calligaris et al. 1995).

Indeed, GATA1 has been shown to form a tricomplex with the retinoblastoma protein (RB1) and E2F2, stalling cell proliferation and driving terminal erythroid differentiation (Kadri et al. 2009). This interaction was abrogated in GATA1s, suggesting an important role for the N-terminus in RB1/E2F binding (Kadri et al. 2009, Klusmann et al. 2010) . Furthermore, internal deletions in the RB1 binding motif of Gata1 have been reported in TAM and ML-DS patients (Toki et al. 2013). The megakaryocytic transcription factor Runx1 has also been shown by co-immunoprecipitation to physically interact with Gata1. In one study is

was proposed that these interactions depended on Gata1 N- and C-terminals but not the zinc finger domain (Elagib et al. 2003). However, this was queried in a later report showing physical interaction of all investigated patient-specific GATA1 mutants with RUNX1 (G. Xu et al. 2006).

Transgenic mouse studies looking at erythroid differentiation indicated that the short isoform may be able to compensate for loss of GATA1 full length, but only when overexpressed several fold above the endogenous level (Shimizu et al. 2001). This contrasted earlier *in vitro* experiments suggesting that the N-terminal domain was dispensable for megakaryocytic or erythroid differentiation (Blobel et al. 1995, J. E. Visvader et al. 1995, Weiss et al. 1997). An important role for the N-terminus in control of fetal liver megakaryocyte proliferation was demonstrated using a Gata1fl/Gata1s rescue assay of primary GATA1 deficient fetal megakaryocyte progenitors (Kuhl et al. 2005) and cultured megakaryocytes (Muntean et al. 2005). Exclusive expression of *Gata1s* in a transgenic knock-in mouse model resulted in a predominantly embryonic phenotype, with transient hyperproliferation and abnormal maturation of a megakaryocyte progenitors and impaired erythropoiesis leading in a few cases to fetal loss (Z. Li et al. 2005). Similar mutations in *GATA1* have been reported in a small number of cases of Diamond Blackfan Anaemia lacking mutations in genes encoding ribosomal subunits (Sankaran et al. 2012) and in a family with a germline *GATA1* variations leading to the exclusive production of GATA1s, who were found to have a moderate-severe macrocytic anaemia, neutropenia and mild impairment in platelet morphology and function, although no defect in absolute platelet count (Hollanda et al. 2006). Significantly however, in the absence of trisomy 21 neither humans nor mice with *Gata1s* mutations develop TAM or ML-DS (Z. Li et al. 2005, Hollanda et al. 2006).

As yet, why *GATA1s* mutations are so common in the setting of T21 and how the effects of *GATA1s* cooperate with T21 genes and/or the T21 environment to cause disease remain poorly understood.

T21 itself results in abnormal fetal haematopoiesis. In DS fetal liver, the immunophenotypic haematopoietic stem cell (HSC) compartment and megakaryocyte-

erythroid progenitor (MEP) population are increased and there is a decrease in B-lymphoid development compared to disomic controls (Roy et al. 2012). The underlying molecular mechanisms driving these changes have been challenging to study *in vivo* - in part because orthologs of human chromosome 21 map to segments of three different mouse chromosomes (Mmu16, 17 and 10). A number of elegant DS mouse models have been generated (reviewed in Herault et al. 2017) . Although none of these models fully recapitulate the haematopoietic abnormalities seen in human T21, they have provided valuable insights into potential contributors to ML-DS pathogenesis (Klusmann et al. 2010, Malinge et al. 2012, Alford et al. 2010, Birger et al. 2013). Ts65Dn mice (trisomic for 104 genes on Mmu16) display a myeloproliferative disorder in adult life (Kirsammer et al. 2008), which has been associated with overexpression of the human chromosome 21-encoded ETS transcription factor *ERG* (Ng et al. 2010). Overexpression of ERG in a mouse model caused expansion of fetal MEPs and double ERG/Gata1s transgenic mice developed a TAM-like condition with increased fetal MEPs, megakaryocyte progenitors and hepatic fibrosis, which progressed to a progenitor myeloid leukaemia (Birger et al. 2013). *DYRK1A* has also been implicated as a mediator of abnormal megakaryopoiesis in ML-DS in a mouse model of ML-DS–like leukaemia, generated by overexpression of consitutitvely active MPL in bone marrow cells from Gata1s/Ts1Rhr mice (trisomic for 33 human chromosome orthologues) by retroviral transduction (Malinge et al. 2012). Roles for mir-125b (Klusmann et al. 2010) and the Insulin-like Growth Factor (IGF) signalling pathway (Klusmann et al. 2010) have also been postulated.

Additional insights have been gleaned using induced pluripotent stem cells (iPSCs) from patients with T21 and GATA1s mutations. IPSCs from patients with TAM mutations were shown to have enhanced megakaryocytic and myeloid potential but impaired erythroid differentiation compared to GATA1 wild type controls (Byrska-Bishop et al. 2015). A separate iPSC study found that T21 led to enhanced production of early haematopoietic progenitors and upregulation of

GATA1s, with altered dosage of Chr21-encoded *RUNX1, ERG* and *ETS2* proposed to play a contributory role (Banno et al. 2016).

The reality may also be more complex than simple dosage adjustment in individual T21-associated genes. Letourneau et al. studied gene expression in fetal fibroblasts from a pair of monozytotic twins – one with T21 and one without. Global transcriptional differences were observed, with domains of genes along all chromosomes either up or down regulated (Letourneau et al. 2014).

1.5 The role of cohesin mutations in transformation to ML-DS

We and others have shown that genes encoding cohesin complex genes are strikingly enriched at transformation from TAM to ML-DS (see chapter 3) (Yoshida et al. 2013, Labuhn et al. 2019).

1.5.1 The cohesin complex

Cohesin is a highly conserved multi-protein complex comprising 4 main subunits: SMC1, SMC3, RAD21 and STAG1/2. These form a ring-like structure around DNA, ensuring cohesion of newly replicated sister chromatids and DNA repair in S-Phase (Nasmyth et al. 2009). This role of cohesin is essential and ensures that genomic information is faithfully passed on through cell division (Figure 1.5).

Cohesin does not bind DNA directly and therefore does not recognise specific DNA motifs. Instead it encircles chromatin fibres. The association between cohesin and chromatin during the cell cycle is dynamic (summarised in Figure 1.6) and involves the coordinated interplay of several cohesin-associated proteins.

Stag1/2 binds to the central region of Rad21 and serves as an interaction site for some of these proteins, including Pds5 and Wapl (which mediate unloading of cohesin from chromatin during the cell cycle) (Losada et al. 2005, Gandhi et al. 2006) and also the transcriptional regulator CTCF (Xiao et al. 2011). During mitosis, cohesin prevents premature separation of sister chromatids and in its absence, chromosome missegregation is common (Remeseiro et al. 2013, White et al.

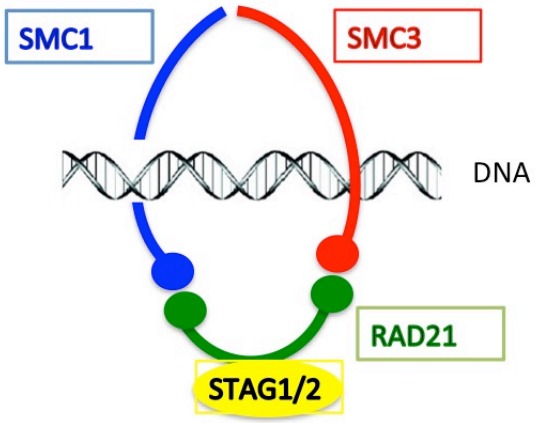


Figure 1.5: The cohesin complex. SMC1A and SMC3 heterodimerise via hinge domains and the ATPase heads associate with RAD21 to form a tripartite ring structure. The STAG subunit (SA1/SA2) interacts with RAD21.

2013). It is postulated that this effect may be dependent on the level of cohesin in the cell, explaining why patients and mice with heterozygous depletion of cohesin components do not display an increased rate of aneuploidy (Remeseiro et al. 2013). Indeed, in one study in yeast, chromosome segregation was only affected when cohesin level was reduced to less than 13% of wild type (Heidinger-Pauli et al. 2010).

1.5.2 A role for cohesin in regulation of gene expression

In higher eukaryotes, cohesin is a major component of chromatin in both non-cycling and post mitotic cells, suggesting an important function outside of the cell cycle (Pauli et al. 2008, Seitan et al. 2011). In particular, there is growing evidence for cohesin's role in the regulation of long-range chromosomal interactions influencing gene expression, in association with CTCF and other transcriptional regulators (Merkenschlager et al. 2013).

It is now widely accepted that the eukaryote genome is not linear, but folded into a three dimensional structure within the nucleus (Figure 1.7). This compaction provides another level at which gene expression can be controlled as distant cis-acting regulatory elements are brought into close proximity with each other

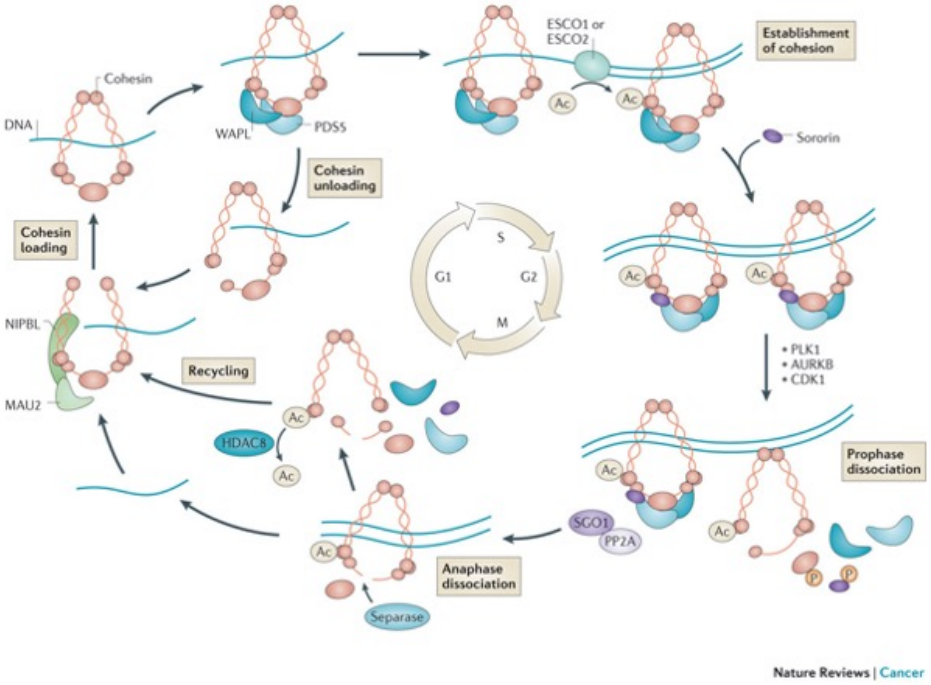


Figure 1.6: Cohesin and the cell cycle. Starting in early G1, cohesin is loaded onto chromatin by cohesin-associated proteins Nipbl and Mau2 and unloaded by Wapl and Pds5A/B in dynamic equilibrium. Cohesin complexes may encircle either a single chromatin fibre or two fibres at the base of a loop bringing distal regions together. During S-phase, acetylation of Smc3 by Esco1 and Esco2 allows interaction with the protein Sororin (which counteracts dissociation activity of Wapl and Pds5A) to allow cohesin complexes to become more stably bound to chromosomes and ensure sister chromatid cohesion. Cohesion is maintained genome-wide throughout G2 when it assists DNA double-strand break repair. During prophase, cohesin is unloaded from chromosome arms by Wapl-Pds5 and chromosome condensation is promoted by another Smc complex, Condensin (not shown). Chromosomes condense, and cohesion is only maintained at the centromeres where Sgo1 and PP2A prevent dissociation. At the onset of anaphase, active Separase cleaves Rad21, allowing chromosome segregation to take place. Reproduced with permission from Losada 2014 Springer Nature.

(Denker et al. 2016). Chromosome conformation capture technologies have revealed that chromosomes can be divided into structural domains termed Topologically associating domains (TADs). TADs contain preferentially interacting regulatory elements and are separated by boundary elements, which may act to prevent inappropriate interactions between enhancers and promoters of neighbouring genes (Merkenschlager et al. 2016, Dixon et al. 2012). The highly conserved zinc-finger

DNA binding protein CTCF is enriched at TAD boundaries and depletion of CTCF has been shown to alter genomic looping and disrupt the TAD structure, suggesting an important regulatory role (Nora et al. 2012).

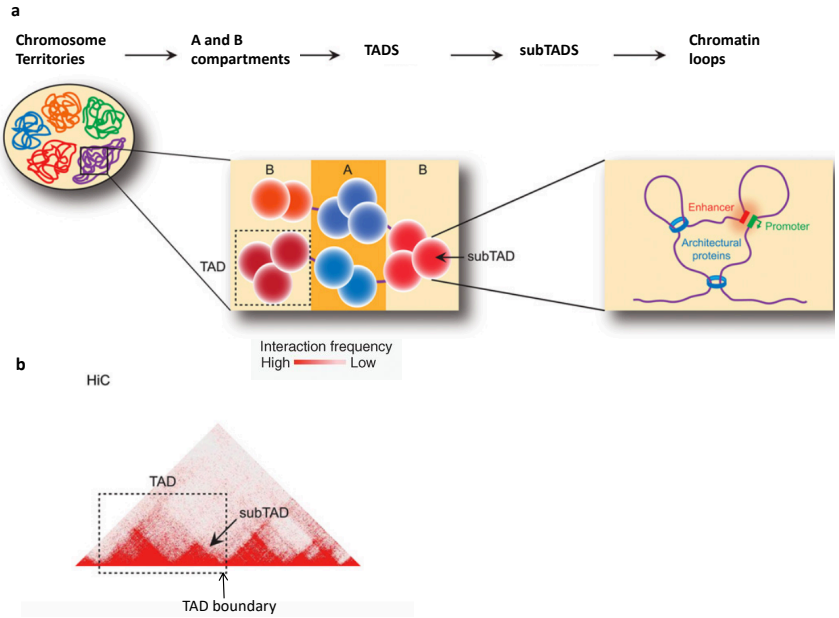


Figure 1.7: Organisation of the genome. Schematic of hierarchical chromatin structure (a). Chromosomes are organised into Topologically associating domains (TADs) which may interact with each other. TADs segregate into A (active shown in blue) and B (inactive shown in red) compartments. They are further divided into smaller contact domains or subTADs in which enhancers, promoters and architectural proteins (e.g. CTCF and cohesin) form chromatin loops. (b) Heatmap from HiC data showing preferential interactions within TADs and subTADs. TAD boundaries restrict interactions between neighbouring TADs. Adapted from Denker et al. 2016, Genes Development.

Cohesin co-localises with CTCF extensively throughout the mammalian genome (Wendt et al. 2008) where it is postulated to regulate interactions between multiple loci within these architectural compartments (Seitan et al. 2013). Depleting cells of CTCF reduces the proportion of cohesin found at CTCF sites, but not the total amount of chromatin-bound cohesin, suggesting that CTCF may influence cohesin distribution rather than cohesin loading (Nora et al. 2012). One suggested model is that, once loaded into chromatin, cohesin extrudes a DNA loop, bringing distant cis-acting elements close together. CTCF-binding sites then act as an barrier, halting loop extrusion and joining the boundaries of a chromatin domain (Merkenschlager et al. 2013).

Postmitotic murine cells depleted of RAD21 show transcriptional dysregulation and loss of chromatin contacts at multiple loci when RAD21 levels are reduced to below 50% of wild type (Sofueva et al. 2013). Similarly, when HEK293T cells were depleted of cohesin through proteolytic cleavage of Rad21, a global loss of short-range enhancer-promoter interactions was seen within TADs (Zuin et al. 2014). Interestingly however, a number of recent reports have suggested that although TAD or sub-TAD structures may be affected by loss of cohesin and/or CTCF, gene expression is often only subtly altered (Nora et al. 2012). In one study (Despang et al. 2019), fusion of TADs at the Sox9-Kcnj2 locus was achieved only following removal of all CTCF sites (both within the TADS and at the boundary). This did not result in any major changes in gene expression, suggesting that, at this locus at least, TADs may be dispensable for gene regulation. By contrast, inverting or inserting new boundary elements resulted in aberrant gene expression and phenotype.

Cohesin also associates with sites independent of CTCF, occupying enhancer and core promoter sites bound by the Mediator complex, which in turn associates and recruits RNA Pol II (Kagey et al. 2010). A number of transcription factors, including Gata1, have been shown to directly interact with the mediator complex (Stumpf et al. 2006). Cohesin is one of the last protein complexes to remain bound to condensing chromatin at mitosis. Therefore, it has been suggested that cohesin may facilitate re-association of transcription factors upon exit from mitosis helping to re-establish and maintain patterns of gene expression through progression of the cell cycle (Yan et al. 2013). The role of cohesin in gene regulation is discussed in more detail in Chapter 5.

1.5.3 Cohesin in normal haematopoiesis

Until relatively recently, little was known about cohesin function in normal haematopoiesis. The human multisystem disorder, Cornelia de Lange Syndrome (CdLS), is an inherited condition caused by heterogeneous mutations in cohesin complex genes (usually *NIPBL*; rarely *SMC1, SMC3* or *RAD21*) (Liu et al. 2009). CdLS cells display no major defects in sister chromatid cohesion but have altered gene

expression profiles. No significant blood count abnormalities are described in CdLS patients (Schrier et al. 2011, Lambert et al. 2011).

Insights into the impact of cohesin gene mutations on haematopoiesis have been gained from animal models. RNAi-mediated depletion of rad21 in zebrafish resulted in loss of expression of runx1 and a failure in mature blood cell differentiation. Microinjection of runx1 mRNA partially rescued the phenotype (Horsfield et al. 2007). While mice homozygous for mutations in the core cohesin components are embryonic lethal, heterozygous knockouts of *Rad21, Smc3* and *Stag1* survive without lethality to adulthood (Remeseiro et al. 2013, White et al. 2013, H. Xu et al. 2010).

It has been shown that haematopoietic heterozygous knockdown of *Rad21, Smc3* and *Stag1* results in increased self-renewal, expansion of the immunophenotypic multipotent myeloid progenitor compartment and impaired differentiation (Mullenders et al. 2015, Viny et al. 2015, Mazumdar et al. 2015). To varying degrees in the different models, cohesin depletion resulted in changed chromatin accessibility. In an *Smc3* conditional knockout, Viny et al. reported a significant decrease in expression of a large subset of genes in cohesin haploinsufficient bone marrow cells compared to wild type, with reduced chromatin accessibility in the -30kb to +10kb region surrounding the transcriptional start site of downregulated genes (Viny et al. 2015). Similarly, Mullenders et al. described increased accessibility of erythroid and myeloid genes in shRNA-mediated cohesin depleted haematopoietic stem and progenitor cells which correlated with changes in gene expression (Mullenders et al. 2015). Finally, in cohesin mutant AML cell lines and in human CD34+ cells, Mazumdar et al. described a global decrease in chromatin accessibility but with increased accessibility at key sites enriched for GATA2, RUNX1 and ERG transcription factor binding. Depletion of these transcription factors reverted the cohesin-mutant phenotype (Mazumdar et al. 2015).

1.5.4 Cohesin gene mutations in myeloid leukaemia

Even in the absence of GATA1s, mutations in cohesin complex genes have been shown to contribute to leukaemogenesis. Recurrent acquired mutations in cohesin

complex genes have been identified at high frequency in solid tumours (Hill et al. 2016). In haematological cancer, mutations are found in 10% of adult AML and myelodysplastic syndrome (MDS) (Kon et al. 2013, Ley 2013, Thol et al. 2014, Thota et al. 2014). Analysis of cohesin-mutated leukaemias based on variant allele frequency (VAF) suggests that these cohesin mutations are present in the dominant clone (Kon et al. 2013; Thol et al. 2014; Thota et al. 2014) and may be relatively early events, although insufficient to drive leukaemogenesis in the absence of additional co-operating mutations. Indeed, in patients with Cornelia de Lange syndrome no significant increase in solid tumours or leukaemia have been described (Lambert et al. 2011, Schrier et al. 2011), although, recently, individual case reports of acute lymphoblastic leukaemia (Fazio et al. 2019) and acute megakaryoblastic leukaemia (Vial et al. 2018) in patients with Cornelia de Lange syndrome have been published.

In both adult AML and ML-DS, the identified cohesin gene mutations are usually mutually exclusive, heterozygous and have a mutational profile suggestive of loss of function. Mutant cells do not display increased aneuploidy or gross chromosomal changes, suggesting the mutations may perturb haematopoiesis through effects on gene transcription rather than the cell cycle.

Understanding how cohesin gene mutations are contributing to the leukaemic phenotype is likely to improve our understanding both of cellular transformation and cohesin biology.

1.6 Summary and Aims

Cohesin gene mutations are strikingly enriched in ML-DS. Why this might be and how cohesin gene mutations co-operate with with Gata1s and T21 in the haematopoietic stem and progenitor compartment to disrupt normal haematopoiesis remains poorly understood. One possibility is that cohesin gene mutations (in the context of megakaryocyte proliferation and differentiation delay driven by T21 and Gata1s) are acting at multiple loci to alter chromatin structure and impact gene regulation, ultimately leading to cellular transformation and leukaemogenesis.

This project focuses on the question of mutational co-operativity between exclusive expression of *Gata1s* and cohesin haploinsufficiency. In order to better understand how these mutations disrupt normal haematopoiesis, both on their own and in combination, I set the following aims:

1. To characterise the type and spectrum of cohesin gene mutations in ML-DS patients.

2. To develop and analyse a mouse model of *Gata1s* and cohesin haploinsufficiency.

3. To identify putative mechanisms by which *Gata1s* and cohesin gene mutations may act together to perturb normal haematopoiesis.

2
Materials and Methods

Contents

2.1 Human studies

2.1.1 Patient sample annotation and processing

Peripheral blood and bone marrow samples were collected from neonates with known or suspected Transient Abnormal Myelopoiesis (TAM) or Myeloid Leukaemia of Down Syndrome (ML-DS). Written informed consent, in accordance with the Declaration of Helsinki, was obtained from parents or custodians of all patients

and the study approved by ethics committees from the Thames Valley Research group, Hannover Medical School and Martin-Luther-University Halle-Wittenberg. Karyotype data was obtained from referring clinicians, the National Down Syndrome Cytogenetic Register, Queen Mary University of London (England) and regional cytogenetic centres (West Scotland and South-east Scotland).

2.1.2 DNA extraction and *GATA1* mutational analysis

Genomic DNA was extracted from patient blood and bone marrow samples using the DNeasy Blood and Tissue Kit (Qiagen), according to manufacturer's instructions.

Patient samples were analysed for exon2/exon3.1 *GATA1* mutations by Sanger sequencing and Direct High Performance Liquid Chromatography (DHPLC) (WAVE; Transgenomic, Omaha, NE). *GATA1* primers were: Exon 2: forward 5-AAAGGAGGAAGAGGAGCAG-3 and reverse 5-AAGCTTCCAGCCATTTCTGA-3'. Amplicon 432 bp and annealing temperature 60°C; Exon 3.1 forward 5GGAACTTGGCCACCATGTTGG-3 and reverse 5-AGCCGCTCTGTCTTCAAAGTCTC-3 (amplicon 310bp), annealing temperature 58°C. PCR conditions were; 5–10 minutes at 95°C, 35 cycles of 1 minute at 95°C, 1 minute at the annealing temperature given and 1 minute at 72°C. After the last cycle, an additional step of 5– 10 minutes at 72°C was performed. PCR products were verified by gel electrophoresis (Alford et al. 2011).

2.1.3 Amplicon re-sequencing and variant calling

Amplicons were designed to cover genes important in ML-DS transformation and myeloid leukaemia. In total 52 TAM, 54 ML-DS and 5 relapse samples were re-sequenced using 1070 150-210 bp amplicons covering a total of 62 genes (appendix A.1). Pooled amplicons were amplified using the 48.48 Fluidigm Access Array, according to manufacturer's recommended conditions and sequenced on the Illumina MiSeq.

Variant analysis was performed on all TAM and ML-DS samples in which *GATA1* mutation(s) could be detected by next generation sequencing. An in-house pipeline

was used for mapping and variant calling (Quek et al. 2016b). FastQ files were trimmed and mapped to the hg19 reference genome using Stampy (Lunter et al. 2011). Bam files were sorted using SAMtools and variant detection performed using VarScan somatic (Koboldt et al. 2013), Pindel (Ye et al., 2009) and GATK (McKenna et al. 2010). VarScan somatic compared TAM or ML-DS sample data to a *GATA1* mutation negative control (remission sample or WT control). VarScan somatic parameters were as follows: min-coverage 100, min-avg-qual 35, min-var-freq 0.05, p-value 0.05. Variants were annotated using Annovar (K. Wang et al. 2010) and, based on this output, filtered by removal of synonymous variants, intergenic regions, known SNPs (dbSNP or 1000 Genomes) (Sherry et al. 2001), highly recurrent calls suggestive of sequencing artefact and previously unannotated SNVs with variant allele frequency of <10%. Remaining variants were manually inspected using the Integrated genome viewer (IGV) and removed if they occurred in duplicated regions, low mapping regions or in areas of poor sequencing quality. Remaining variants were retained if they were previously documented as oncogenic in the COSMIC database (Tate et al. 2019), were present at 5x greater than background when manually assessed using the Grep function to specifically search for the variant or had been detected in a previous sample from the same patient.

2.2 Mouse studies

2.2.1 Breeding and backcrossing of mouse lines

All mice were bred and maintained in accordance with UK Home Office regulations.

Conditional *Rad21*$^{fl/+}$ mice (Seitan et al. 2011) were a gift from the Nasmyth lab. These mice were backcrossed at least 6 generations onto a C57BL/6 background and bred with *VavCre* tg mice, also on a C57BL/6 background (Boer et al. 2003) to facilitate haematopoietic-specific knockdown of *Rad21*. Backcrossing was performed using a speed congenic approach, with background confirmed by sending DNA to Charles River Laboratories for comparison to a 384 single nucleotide polymorphisim (SNP) panel typical of C57BL/6.

Gata1s (ATFG1s-2H7B/BirA(Rosa26)) mutant mice were developed in the Vyas lab (G.Juban, unpublished). In this model, the truncated Gata1 protein is tagged with a biotin tag, TEV cleavage site and flag tag. These mice require BirA ligase to facilitate biotinylation of the tag. They were originally on a mixed background (120ola, C57BL/6, CD-1, FVB/N) and were backcrossed onto a C57BL/6 background. These mice were then crossed with *Rad21*$^{fl/+}$ *VavCre* tg mutants to create combined *Gata1s Rad21*$^{fl/+}$ *VavCre* tg mutants.

True wild type, *VavCre* tg only or *Rad21*$^{fl/+}$ *VavCre* WT were used as wild type controls. Unless specified, analysis was performed in 6-12 week old male mice. Where possible mice were analysed together with littermate controls of the relevant genotype.

The alternative *Gata1s* model (*Gata1*^{efl}) was a gift from the Orkin lab. These mice were on a mixed background and were genotyped using the primers and parameters described (Z. Li et al. 2005).

2.2.2 Genotyping of mice

Ear punch tissue samples from mice were lysed in 100uL lysis buffer (Tris-HCL 1M 2.5 mL, EDTA 0.5M 0.1 mL, Tween-20 0.25 mL made up to 50mL with water) and proteinase K for 1-2 hours or overnight at 37°C. The lysate was directly used as template for PCR reactions.

VavCre: Primers Vav-iCre F (5' – AGATGCCAGGACATCAGGAACCTG – 3') and Vav-iCre R (5' – ATCAGCCACACCAGACACAGAGATC) using the following parameters: 95 C for 1min followed by 29 cycles of 95°C for 10s, 60°C for 10s and 72°C for 10 sec. The VavCre tg mutant allele was detected as a band of 250bp. In a VavCre wildtype sample no band is seen.

Rad21: Primers 6F (5'-TGGAGAAGGAAATGATGGCGG-3') and 6R (5'-ATGGTGGCTCACAACCACTC-3') using the following parameters: 95°C for 1min followed by 29 cycles of 95°C for 10s, 60°C for 10s and 72°C for 10 sec. The wildtype allele was detected as a band of 500bp and the floxed allele as a band of 600bp.

Cre-mediated excision in bone marrow and spleen was confirmed by PCR with primers 4F (5'–GAAGGTCATGGTTGGCAGAT–3') and 6R (5'-ATGGTGGCTCACAACCACTC-3') and the following parameters: 95°C for 1min followed by 29 cycles of 95°C for 10s, 60°C for 10s and 72°C for 10 sec. In the excised allele, a band was detected at 300bp. For completeness, excision of exons 5 and 6 was confirmed on one occasion by Sanger Sequencing of this PCR product.

Gata1s: Primers ATFG F (5'-TCACAGGTTCAACCCCAGTG-3') and ATFG R (5'-GTTGAGGCAGGGTAGAGTGC-3') and the following parameters 94°C for 3 min followed by 30 cycles of 94°C for 30s, 60°C for 30s, 72°C for 50s and then 72°C for 2 min. The wildtype allele was detected as a band of 802bp. The mutant allele (*Gata1s*) was detected as a band of 213bp.

BirA: Primers for mutant: 9414 (5' – TTCAGACACTGCGTGACT -3') and 9415 (5'–GGCTCCAATGACTATTTGC-3'). Primers for wildtype: 9416 (5'–GTGTAACTGTGGACAGAGGAG-3') and 9417 (5'–GAACTTGATGTGTAGACCAGG-3'). The following parameters were used: 94°C for 3 min followed by 34 cycles of 94°C for 30s, 62°C for 30s, 72°C for 30s and then 72°C for 2 min. The WT allele was detected as a band at 367bp. The mutant allele was detected as a band at 500bp.

2.2.3 Peripheral blood analysis

Blood was collected either by tail vein sampling or, at the time mice were sacrificed, by direct cardiac puncture. Automated peripheral blood counts were obtained using a Sysmex KX-21N 3-part differential haematology analyser. Peripheral blood smears were made and stained using the May-Grünwald/Giemsa-staining method.

2.2.4 Immunophenotypic analysis and sorting of bone marrow

Freshly dissected femurs, tibias +/- cristae and vertebrae were crushed using a pestle and mortar and flushed through a 30uM filter with 10ml of media (5% Fetal

calf serum (FCS) in PBS). Samples were quantitated using the Sysmex KX-21N automated haematology analyser to calculate cellularity.

For analysis, 10 million cells per mouse were used for the myeloid progenitor panel and 15 million cells per mouse for the HSC panel. Cells were incubated with Fc-block and stained with rat anti-mouse antibodies. For the myeloid progenitor panel, Fc-block was not used as this interferes with CD16/32 staining. Fluorescence-minus-one (FMO) and single stain controls were included for all markers. Cells and antibodies were incubated at 4°C for 20 minutes in the dark, washed and then re-suspended in PBS+5%FCS prior to analysis.

Antibodies used for flow cytometery analysis are summarised in Figure 2.1 and listed in full in Appendix A.2. All antibodies were purchased from BD, ebioscience or BioLegend and used at previously titrated concentrations. Samples were analysed using a LSRFortessa X20. Gates were set using Fluorescence minus-one (FMO) controls and, where possible, negative populations.

Cells used in sorting experiments were bead-enriched for Kit using a CD117 bead enrichment kit and MACS cell separation columns (Miltenyi Biotec) prior to staining. Cells were sorted with a BD FACSAriaII instrument using an 85micron nozzle. Sorting purity was assessed before and (provided sufficient cells were available) at the end of sorting, aiming for a purity of >95%. Analysis was performed using FACSDiva v8.1 (BD Biosciences) and Flowjo v10.1 Software.

	HSC Panel			Myeloid Progenitor Panel (Bryder and MkP)	
Lineage Markers	CD5 CD4 CD8a B220 Gr1 TER119	All on PECy5		CD5 CD4 CD8a B220 Gr1	All on PECy5
Antigen, Conjugate and Clone	CD48 c-kit Sca-1 Flt3 CD150	APC APC-eF780 PB PE PECy7		CD16/32 CD150 c-kit Sca-1 CD41 CD9 CD105-Bio	PE APC APC-eF780 PB PECy7 BV650 SA-PETxR
Viability	7AAD			7AAD	

Figure 2.1: Antibodies used for flow cytometric sorting and analysis of mouse bone marrow.

2.2.5 Protein extraction and Western blotting

Bone marrow (from legs, cristae and vertebrae) and spleen cells were harvested from 6-12 week old mice, filtered and incubated for 5 minutes with ammonium chloride buffer to lyse red cells. The remaining cells were then washed with PBS, re-suspended in 10x volume 0.5% triton/RSB 100 with protease inhibitor, sonicated (10x 30seconds on/off) and incubated with benzonaze at 4°C for 1 hour. The sample was quantitated using the Bradford Index. Protein extract was then boiled for 10 minutes in 4x NUPAGE LDS sample buffer with DTT, loaded onto a 4-12% NUPAGE BisTris 15 well gel and run for 1 hour at 150V in NUPAGE MOPS running buffer. Proteins were transferred onto an activated PVDF membrane (Amersham) at 4°C for 2 hours at 3OV in transfer buffer (5% methanol) and blocked in 5% milk/PBS overnight. The membrane was then incubated with 1:1000 in 5% milk/PBS anti mouse Rad21 antibody (992, AbCam) or 1: 5000 anti mouse GAPDH antibody for 1 hour, washed and incubated with 1:10,000 donkey anti-rabbit secondary antibody. The washed membrane was developed using the enhanced chemoluminescence system (Amersham) and Kodak Biomax film.

2.2.6 *In vitro* analysis of sorted cells

Immunophenotypically defined populations were sorted from freshly prepared mouse bone marrow directly into 100uL Iscove's modified Dulbecco's medium supplemented with 10% fetal calf serum (FCS), l-glutamine (Invitrogen), 50 U/ml penicillin, and 50 µg/ml streptomycin (Cambrex) and the following cytokines: thrombopoeitin (TPO) 40ng/ml (Peprotech), erythropoietin (EPO) 2U/ml (Roche), murine IL-3 10ng/ml (Peprotech), human IL-11 5ng/ml (Peprotech), murine stem cell factor (SCF) 50ng/ml (Peprotech), murine granulocyte-macrophage colony-stimulating factor (GM-CSF) 5ng/ml (Peprotech) and human Flt3-Ligand 10ng/ml (RD systems). Cells were imaged at day 3. On day 5, cells were harvested, counted and labelled with CD41-PE, CD42b-FITC, Kit-APC, TER119-PECy7, Hoechst 33342 and 7-AAD (viability). Details of antibodies are included in Appendix A.2. Unstained, single stained and Fluorescence Minus One (FMO) controls were used to determine

the background and set the compensation in each channel. Data were analysed using FlowJo v10.1 software.

2.2.7 Histopathological analysis of bone marrow

For analysis of the bone marrow, femurs from 6-12 week old male mice were cleaned of soft tissue and fixed in 10% neutral buffered formalin solution for 24 hours followed by gentle decalcification in 10% EDTA in distilled water for up to 48 hours. They were then processed for analysis by the histopathology team at The Kennedy Institute of Rheumatology, University of Oxford, who also performed the hematoxylin and eosin (H+E) staining. Megakaryocytes were quantitated using ImageJ software (v1.52).

2.3 Molecular analysis

2.3.1 RNA-sequencing

For bulk RNA-sequencing, 100 sorted MkP cells (Live, lin-, Sca1-, CD16/32low, CD41+, Kit+, CD9+) were sorted directly into lysis buffer containing RNAse inhibitor (Clontech). External RNA Controls Consortium (ERCC)-RNA spike-in mix 1 (Ambion, Life technologies) was added prior to c-DNA synthesis as per manufacturer's instructions assuming 100pg RNA per 100 cell sample. cDNA synthesis was performed using SmartSeq-V4 Ultra low input RNA kit (Clontech). Illumina libraries were generated using the Nextera XT DNA sample preparation kit and Index Kit (Illumina UK). Library quality and size were checked using the Agilent Bioanalyser (Agilent Technologies UK). Libraries (4nM) were sequenced on an Illumina HiSeq 4000 (Sanger Centre, Oxford) using 75 bp paired end reads.

2.3.2 ATAC-sequencing

ATAC-seq was performed on sorted MkP using a protocol adapted from Corces 2015 with modifications for lower cell numbers. Samples were prepared and analysed in at least biological triplicate. For optimisation, 5000-10,000 freshly prepared cells were sorted into FACS buffer (PBS with 10% FCS), pelleted by centrifugation at 500g

for 10 min and the supernatant was carefully removed. The pellet was resuspended in 50uL lysis/transposition mix (25uL 2xTD buffer, 2.5uL Tn5 transposase, 0.5uL 1% digitonin and 22uL nuclease free water) and incubated at 37°C for 30 min with agitation at 300rpm. The reaction was stopped using Qiagen MinElute Reaction Cleanup kit and the samples eluted in 10uL warmed EB (10mM Tris-HCL, pH8). Samples were then amplified and indexed using NEBNext 2X Mastermix (New England Biolabs) and custom primers as described (Buenrostro et al. 2013) (72°C 5 min, 98°C 30 s, 11 cycles of 98°C 10 s, 63°C 30s, 72°C 1min). Library quality was assessed using a DNA 1000 high sensitivity Screen Tape (Agilent) and quantitated by qPCR using the NEBNext library quantitation kit. ATAC-seq libraries (6nM) were sequenced on the NextSeq 500 platform (Illumina) using a NextSeq High Output v2 75 cycle sequencing kit.

For lower cell numbers (1000-5000 cells), cells were sorted directly into 25uL of lysis/transposition mix (on ice) and then processed as above. Where used, DNAase and EDTA were omitted from the FACS buffer.

2.4 Bioinformatic analysis of sequencing data

2.4.1 RNA-seq data

RNA-sequencing data was quality assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired end reads were trimmed for Nextera transposase sequences using Trimmomatic (Bolger et al. 2014) and mapped to the mm10 mouse genome build using STAR (Dobin et al. 2013). Mapped reads were quantified using featureCounts (Liao et al. 2014) and normalised to Transcripts per Million (TPM). Differential gene expression was carried out in DESEQ2 (Love et al. 2014) in R 3.2.1 (R Core Team 2014). Genes with less than 5 counts across all samples were discarded. Differentially expressed genes were identified using Wald test pairwise analysis. Unless otherwise specified, a p-adj < 0.05 cutoff was applied to obtain significantly up- or down-regulated genes. Dispersion factors were estimated using a local fit and gene counts were normalised using rlog transformation. Plots were generated using ggplot2 (Wickham 2016) and

pheatmap (https://cran.r-project.org/package=pheatmap). The PCA loadings plot was generated using factoextra (https://CRAN.R-project.org/package=factoextra). Enrichment analysis was carried out using EnrichR (Chen et al. 2013) and Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005).

2.4.2 ATAC-seq data

Analysis was performed using a pipeline generated by Jelena Telenius (http://userweb.molbiol.ox.ac.uk/public/telenius/PipeSite.html) with default parameters (Telenius et al. 2018). In brief, Fastq files were aligned to mm9, mm10 or hg19 with Bowtie (Langmead 2010). Unmapped reads were trimmed for adapters using trim-galore and trimmed reads remapped. PCR duplicates and mitochondrial reads were filtered using Samtools (H. Li et al. 2009). The resulting bam file was converted to BigWig format and viewed in the UCSC data hub (Kent et al. 2002) alongside ENCODE published DNase hypersensitivity data (Consortium 2012). For peak calling, bam files from each genotype were merged and peakcalling was performed using MACS2 (Zhang et al. 2008). Peak reproducibility, correlation with published data and distance from transcriptional start sites was investigated using deepTools (Ramírez et al. 2014). Peaks were quantitated using featureCounts (Liao et al. 2014) and differential accessibility analysis was performed using DESEQ2 (Love et al. 2014). Peak annotation was performed using Homer annotatePeaks.pl (Heinz et al. 2010) and ChIPpeakAnno (Zhu et al. 2010) with the promoter region being defined as 1kb up or downstream of the transcriptional start site. EnrichR was used for enrichment analysis (Chen et al. 2013). Motif calling was performed using Homer findMotifsGenome.pl (Heinz et al. 2010) using both MACS2 narrowPeak and MACS2 summit files extended +/- 200bp as input files and a file of all ATAC peaks as background.

2.4.3 Published datasets

Published datasets were downloaded from the NCBI Gene Expression Omnibus (GEO) database (Barrett et al. 2012) and, where necessary, converted to FASTQ files

using the SRAtoolkit (Leinonen et al. 2011) for processing as previously described.

2.4.4 Statistical Analysis

Data were analysed and graphs generated using Prism GraphPad for Mac (v7.0) and Excel (Microsoft). Descriptive statistics show mean ± SEM, unless otherwise stated. Differences in characteristics between groups were assessed using One-way ANOVA or Student's t-test.

3

Cohesin gene mutations in myeloid leukaemia of Down Syndrome

Contents

Work in this section was done in collaboration with Dr Kelly Perkins, Marlen Metzner and Alison Kennedy (Vyas lab, MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford) and forms part of a recent publication in Cancer Cell (Labhun et al. 2019) on which I am an author. Exome and bait sequencing was completed prior to my involvement in the project. Figure 3.4 is adapted directly from this publication.

3.1 Introduction

Like many cancers, ML-DS results from the step-wise acquisition of somatic mutations in stem and progenitor cells. As discussed in Chapter 1.3.2, transformation from pre-leukaemia (TAM) to leukaemia (ML-DS) requires both the persistence of a *GATA1s* mutation and gain of additional co-operating mutations in GATA1s mutant clone(s).

It has been shown that some of these transforming mutations occur in genes encoding signalling molecules, epigenetic regulators and members of the cohesin complex. A role for activating mutations in the tyrosine kinase *JAK3* in ML-DS pathogenesis was first demonstrated by Walters et al. in the acute megakaryoblastic leukaemia (AMKL) cell line CMK and subsequently confirmed in patient samples. These mutations were sufficient to transform Ba/F3 cells to factor-independent growth and induce splenomegaly, expansion of the megakaryocyte-erythroid progenitor populations and impaired megakaryocyte function in a mouse model, but did not cause leukaemia in the absence of trisomy 21 (Walters et al. 2006). Nikolaev et al. subsequently performed exome sequencing of 5 TAM, 3 ML-DS and 2 complete remission samples (Nikolaev et al. 2013). Truncating *GATA1* mutations were invariably found in both TAM and ML-DS cases. Paucity of additional somatic variants in TAM samples suggested that *GATA1s* in the context of T21 was both necessary and sufficient for pre-leukaemia. By contrast, additional potential driver mutations were identified in in ML-DS samples in genes including *EZH2, JAK1, EXT1, DLEC1* and APC.

In a larger study, Yoshida et al. reported deep sequencing of 41 TAM, 49 ML-DS (DS-AMKL) and 19 non-DS AMKL patient samples, including 15 TAM and 14 ML-DS samples which were subject to either whole exome or whole genome sequencing (Yoshida et al. 2013). Once again, the presence of constitutive Trisomy 21 plus a single *GATA1* mutation appeared sufficient for the development of TAM. ML-DS was shown to evolve from pre-existing TAM clone(s) through somatic acquisition of additional mutations. Most commonly, these included components of the cohesin complex (53% of cases) and the closely-associated transcriptional regulator CTCF (20%). Epigenetic regulators, including *EZH2* and *KANSL1* were mutated in 45% of ML-DS cases and *JAK, MPL, SH2B3* and *RAS* pathway genes in 47% of cases. ML-DS samples were compared to non-DS AMKL samples, in which recurrent CBFA2T3-GLIS and RBM15-MKL gene fusions are enriched

(Gruber et al. 2012, Thiollier et al. 2012). No such CBFA2T3-GLIS2 or RBM15-MKL fusions were detected by RT-PCR or RNA-sequencing in ML-DS samples, suggesting different pathways of pathogenesis.

3.2 Aims

Drawing on a large cohort of patient samples, we aimed to build on this work - characterising the mutational landscape of TAM and ML-DS and considering the nature and function of some of the identified transforming variants in more detail.

Given the striking enrichment of cohesin and CTCF gene mutations in ML-DS (compared to both TAM and non-DS AMKL) (Yoshida et al. 2013), I was particularly interested to study how these mutations might co-operate with Gata1s to perturb haematopoiesis and drive leukaemia development.

Therefore, the aims of this first section of work were:

1. To characterise the spectrum of somatic mutations in a large cohort of TAM and ML-DS patients.

2. To identify and describe functionally significant cohesin gene mutations in this cohort.

3.3 Results

3.3.1 Patient Samples

In total, 252 blood or bone marrow samples were included in this study, representing 141 ML-DS and 111 TAM samples from 128 ML-DS and 103 TAM patients. In 12 cases we were able to analyse paired TAM and ML-DS samples from the same patient. 5 of the ML-DS samples represented disease relapse. Where available, remission/post treatment *GATA1* negative samples were included as controls.

For the purposes of analysis, TAM samples were defined as those positive for a *GATA1* variant by either Sanger sequencing and/or next generation sequencing, collected within 3 months of birth and from patients who subsequently demonstrated clinical and/or molecular resolution. ML-DS samples were defined as being positive for a *GATA1* variant, collected greater than 3 months from birth and/or showing clinical features of leukaemia. ML-DS relapse was defined if a patient was post-treatment and *GATA1* mutation positive.

A summary of the analysed samples is shown in Figure 3.1. As might be expected, the majority of TAM samples were peripheral blood samples, while the majority of ML-DS samples were bone marrow. Mean blast % was equivalent in TAM and ML-DS, but increased in the small group of relapse samples. In TAM the majority of patients were shown to have no additional cytogenetic abnormalities other than T21. The number of additional karyotypic changes noted was increased in ML-DS and relapse samples.

	TAM	ML-DS	Relapse
Total number	111	136	5
Male	64	75	3
Female	47	61	2
Peripheral blood	106 1 fetal blood	23	0
Bone marrow	4	113	5
Mean blast %	31% (0-94)	35% (0-95)	47.5% (10-90)
Karyotype			
+21 only	99	70	0
Additional abnormalities	6	46	3
Not available	6	20	2

Figure 3.1: Summary of sample dermographics.

3.3.2 Next generation sequencing and variant calling

Prior to my involvement in the study, 7 TAM and 13 ML-DS samples were analysed by whole exome sequencing. An additional 86 TAM and 107 ML-DS samples were re-sequenced using RNA bait analysis covering 111 genes and additional chromosomal changes (Papaemmanuil et al. 2013). Working with Dr Kelly Perkins (post-doc, Vyas Lab), I designed amplicons and performed amplicon-based sequencing of a further 53 TAM and 58 ML-DS samples covering 62 genes. 67 samples were sequenced by more than one method, allowing internal validation of results. Amplicons were designed to cover variants previously reported in ML-DS patients (Yoshida et al. 2013, Nikolaev et al. 2013), cohesin complex genes and additional genes important in myeloid disease. In total, 1070 amplicons were needed to achieve adequate coverage of 62 genes (Appendix A.1). A summary of the genes covered by the amplicon panel is shown in Figure 3.2 and average sequencing depth over each gene shown in Figure 3.3. Only one amplicon (ADRA2C) failed.

Variant analysis was carried out using an in-house mapping and variant-calling pipeline (Quek et al. 2016a). Two variant callers were used: VarScan somatic (Koboldt et al. 2013) to detect single nucleotide variants and small indels compared to *GATA1* variant negative control samples, and Pindel (Ye et al., 2009) which has slightly better detection of larger indels. The variant filtering criteria applied are described in detail in Materials and Methods.

3.3.3 ML-DS is enriched for additional somatic variants

In keeping with published reports, at least one N-terminal truncating *GATA1* mutation was detected in all TAM and ML-DS samples. ML-DS was significantly enriched for additional single nucleotide variants (SNVs) and small insertion/deletions (indels) compared to TAM - with 1.6 variants detected per sample in ML-DS and 0.4 variants per sample in TAM ($p = 0.0000001$, two-tailed student T-test). The most commonly acquired somatic variants in ML-DS were in cohesin complex genes and *CTCF* (47% of samples), signalling molecules (*JAK, MPL* and *KIT*) (48% of samples)

Category	Gene
Gata1	GATA1
Cohesin	STAG2
	RAD21
	SMC1A
	SMC3
	STAG1
	CTCF
	NIBPL
Epigenetic	EZH2
	BCOR
	KANSL1
	SUZ12
	KDM6A
	KMT2A
	KMT2C
	KMT2D
	EP300
	KAT8
	TET2
	DNMT3A
	PHF8
Tyrosine kinase	DCAF7
	JAK1
	JAK2
	JAK3
	MPL
	SH2B3
	PDGFRB
	FLT3
	GNB1
	PTEN
	PTPN11
	KIT
	CBL
	PTPRD

Category	Gene
RAS	NRAS
	KRAS
	NF1
Transcription Factor	ATRX
	RUNX1
	ASXL1
	CREBBP
	CUX1
	MYC
Other	CSF2RB
	SRSF2
	ADRA2C
	CSF2
	DLEC1
	TGD
	ZRSR2
	NOTCH2
	EXT1
	MN1
	GFPT2
	IL3
	ITGB1
	SETBP1
	IDH1
	IDH2
	SF3B1
	WT1

Figure 3.2: Genes covered by amplicon resequencing panel.

and epigenetic regulators (14% of samples) (Figure 3.4). Additionally, in 6 ML-DS cases, a novel hotspot mutation was detected in the *CSF2RB* gene, which encodes the common beta chain of the IL3, IL5 and GM-CSF receptors. This was subsequently shown by collaborators in the Constaninescu lab to be functionally oncogenic (Labuhn et al. 2019).

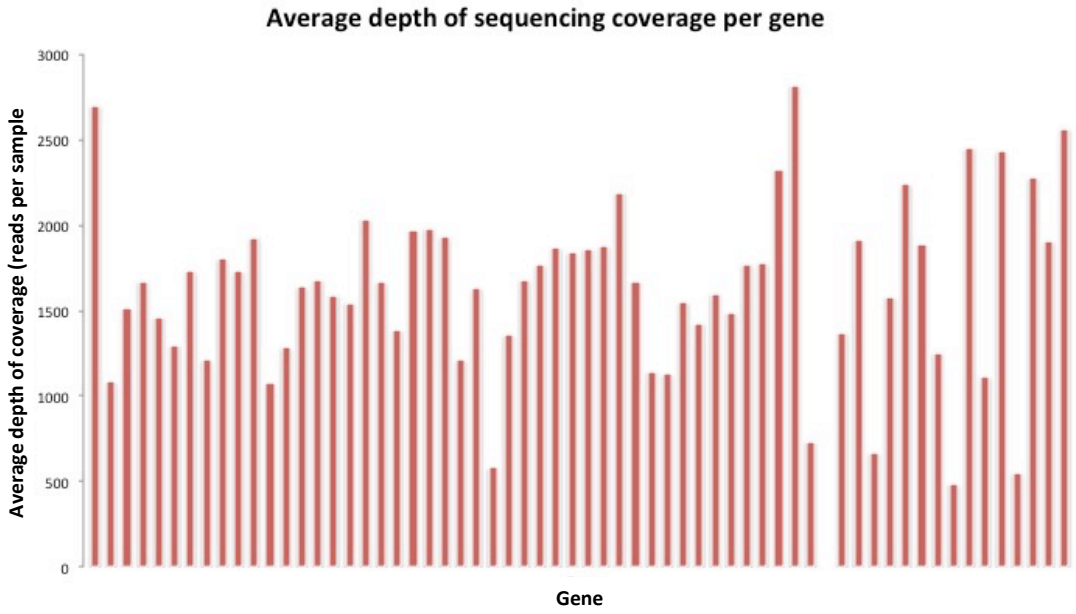


Figure 3.3: Average depth of coverage for genes covered by the amplicon panel. Y-axis represents average number of reads per sample for each gene (taking into account all overlapping amplicons). Each column represents a gene.

3.3.4 The functional impact of detected variants

The majority of ML-DS cohesin and epigenetic regulator variants were frameshift, nonsense or large deletions and therefore predicted to result in loss of function (shaded red in Figure 3.4). Although occasional variants were detected at these loci in TAM samples, these were mainly missense SNVs and less likely to have a deleterious effect (shaded blue in Figure 3.4). The *JAK2, JAK3* and *MPL* variants in ML-DS were classified as likely gain of function variants based on correlation with previous reports (green boxes in Figure 3.4). In TAM, however, 3/5 *JAK* variants were predicted to result in loss of function and the remainder were previously undescribed SNVs. Subsequent confirmatory functional experimental work (Labuhn et al. 2019) revealed that none of the *JAK* or *MPL* variants reported in TAM samples resulted in downstream JAK-STAT signalling activation.

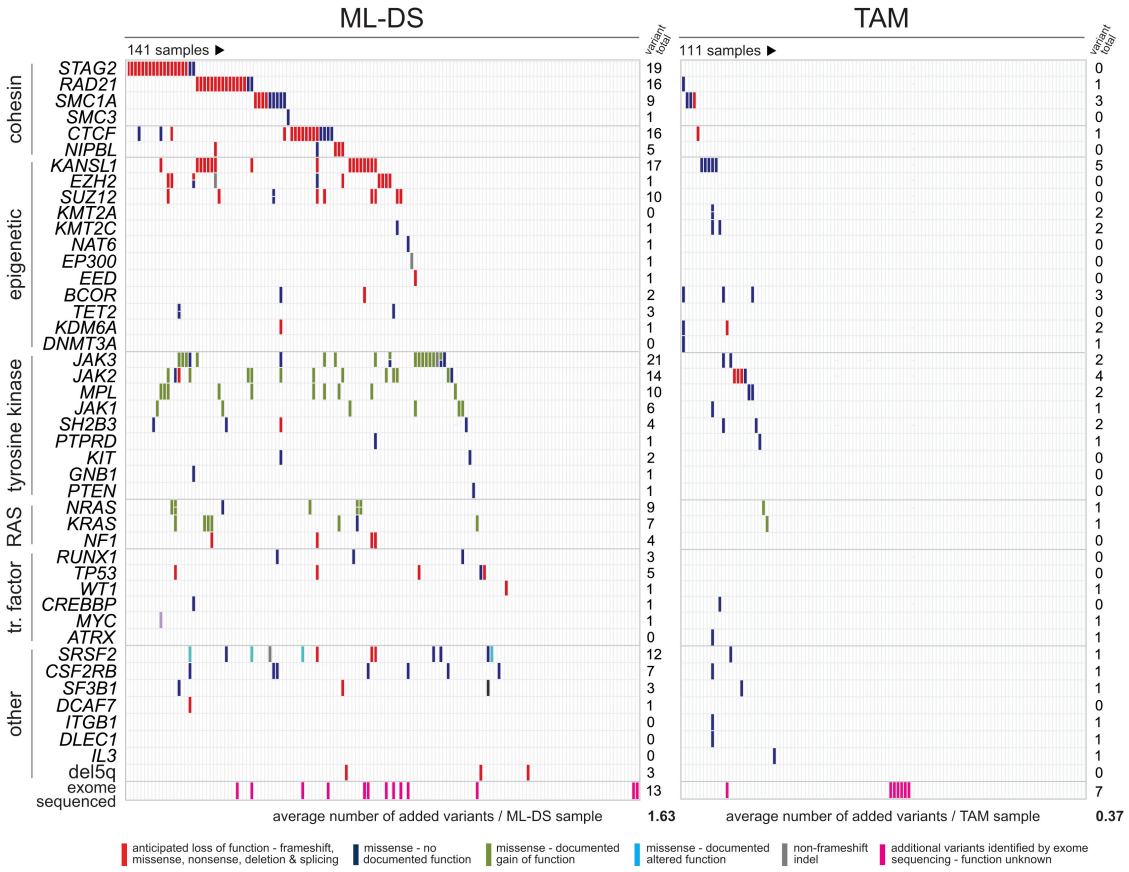


Figure 3.4: Mutational landscape of ML-DS and TAM. Summary of data for ML-DS (left panel) and TAM (right panel) sequenced samples. Each column represents a sample and each row represents a gene. Shaded boxes indicated a detected variant, with predicted functional impact colour-coded according to the legend at the bottom of the figure. Adapted from Labuhn et al., 2019 with permission from Elsevier.

Interestingly, in 25% of the ML-DS samples, no somatic variants other than *GATA1s* were detected. In an attempt to address this, I performed RNA-sequencing on samples where sufficient material was available to look for fusion transcripts, but none could be identified. Possible explanations for this sample set are the presence of lower VAF variants which were not detected as significant using the current pipeline settings, variants not covered by the 62 gene sequencing panel used and/or additional mechanisms of leukaemogenesis.

3.3.5 Cohesin gene mutations in ML-DS

I next examined the identified cohesin variants in greater detail. As shown in Figure 3.4, cohesin gene mutations were amongst the most commonly identified

variants in ML-DS samples. With a few exceptions, these were mutually exclusive and heterozygous. A number of ML-DS samples were identified where only T21, a *GATA1s* mutation and a cohesin gene mutation were identified in the sample, suggesting that in the right cell context this combination of factors may be both necessary and sufficient for leukaemia. The relative numbers of cohesin variants across ML-DS is summarised in Figure 3.5a. Similar to the findings of Yoshida et al. (2013), *STAG2* and *RAD21* were the most commonly mutated cohesin loci, with variants also detected in *SMC1A, SMC3, NIPBL* and the associated transcriptional regulator *CTCF*. The majority of *RAD21* variants were nonsense, frameshift or large deletions (Figure 3.5b) and therefore strongly suggestive of loss of function. *STAG2* mutations followed a similar pattern, although a number of splice-site variants were also detected. By contrast, the majority of variants detected in *SMC1A* were missense variants, and therefore more challenging to functionally characterise.

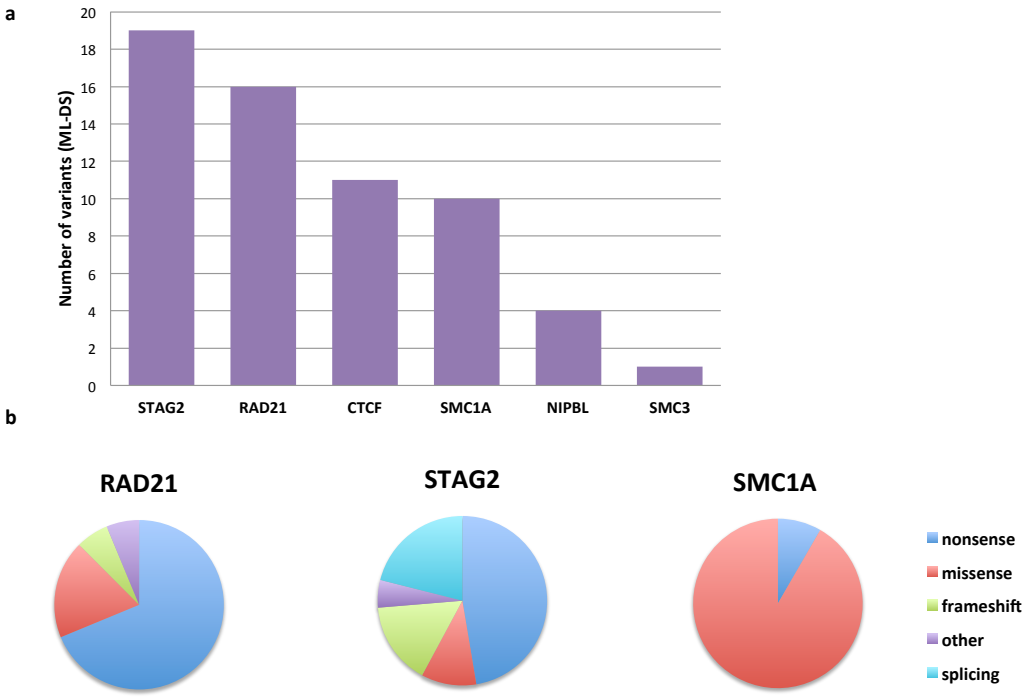


Figure 3.5: Cohesin gene mutations in ML-DS. (a) Histogram showing number of cohesin variants detected in ML-DS samples. (b) Classification of detected RAD21, STAG2 and SMC1A variants. Data from Labuhn et al. 2019.

RAD21 is a core cohesin complex component and potentially a good target for further studies: (a) It is commonly mutated in ML-DS, and can be the only variant in addition to T21 and *Gata1s*; (b) The majority of detected variants are nonssense or frameshift and therefore predicted to result in loss of RAD21 function; and (c) Unlike STAG1/STAG2 and SMC1A/SMC3, where there has been the suggestion of a degree of functional redundancy between partners (although recent reports suggest that this may be complex (Viny et al. 2019)), RAD21 does not co-exist with a structurally similar partner.

Variants found in *RAD21* were scattered throughout the gene (Figure 3.6), although there was the suggestion of some clustering at *SMC3* and *SMC1A* binding sites where even small variations might result in disruption of the cohesin ring. In Figure 3.6, which schematically depicts the *RAD21* locus, variants detected in our study (red) are shown alongside those reported by Yoshida et al. (2013) (green) and the one missense variant detected in a TAM sample (grey).

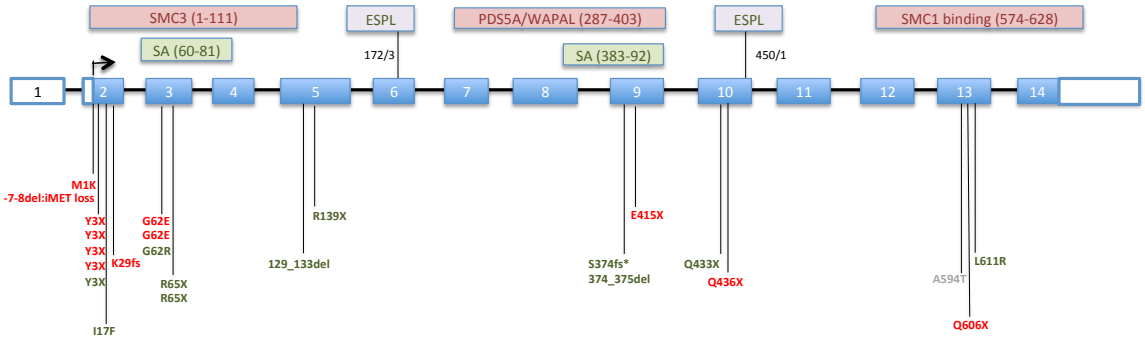


Figure 3.6: Schematic of amino acid positions of *RAD21* variants detected in ML-DS. Exons depicted as shaded blue/white boxes. Black horizontal lines represent introns. Shaded pink and green boxes show binding sites for SMC3, SMC1A, STAG (SA) and PDS5a/WAPL. Rad21 cleavages sites are marked as "ESPL" in grey boxes. Variants are annotated underneath the locus. Variants in red are from our dataset and green from Yoshida et al. (2013).

3.4 Discussion

These data, from a large cohort of TAM and ML-DS patients, highlight a number of important points: (1) Compared to the pre-leukaemic TAM state, ML-DS is characterised by additional somatic variants. These include genes encoding signalling molecules, epigenetic regulators and members of the cohesin complex. Findings are consistent with those of other groups (Yoshida et al. 2013, Nikolaev et al. 2013) and comprise a relevant resource for future studies; (2) Using predictions based on the nature of mutations and/or annotation against databases such as COSMIC, it was shown that the majority of variants seen in ML-DS are likely to result in loss (cohesin complex components and epigenetic modifiers) or gain of function (JAK signalling pathways), while the few somatic variants reported in TAM samples are mainly missense and less likely to be functionally relevant, as shown by studies in the Constantinescu lab (Labuhn et al. 2019); (3) A striking enrichment for mutations in cohesin complex components (*RAD21, STAG2, SMC1A, SMC3* and the associated *CTCF* and *NIPBL*) was seen in ML-DS compared to TAM, molecular remission and non-DS AMKL. *STAG2* and *RAD21* were the most commonly mutated cohesin complex components and variants in these genes were strongly predicted to result in loss of protein function and ineffective formation of the cohesin ring; (4) As seen previously in both ML-DS (Yoshida et al. 2013) and cohesin mutated adult AML (Kon et al. 2013, Ley, 2013, Thol et al. 2014; Thota et al. 2014), cohesin gene mutations were largely mutually exclusive, suggesting that overall loss of cohesin function was the dominant effect. With the exception of *SMC1A* and *STAG2*, which are located on the X chromosome and therefore hemizygous in male patients, mutations were predicted to be heterozygous based on VAF analysis. This supports the assumption that a certain level of cohesin is required for cell division and survival. There is some suggestion that SMC1A and SMC3 and STAG1/STAG2 may provide a degree of functional compensation for each other although recent studies have also supported unique roles for these proteins (Viny et al. 2019). Significantly, it would seem that, at least in some cases, T21, GATA1s and a cohesin gene mutation are sufficient for ML-DS.

A number of interesting questions are raised by these findings. Why are *GATA1* mutations so common in the context of T21? How do T21 and GATA1s interact to cause disease? Why do mutations in the genes encoding the cohesin complex occur at a higher frequency in ML-DS than adult AML? How do T21, *GATA1s* and cohesin gene mutations co-operate to perturb haematopoiesis and cause leukaemogenesis?

Picking apart this final question is relevant but challenging - particularly with regards to addressing the molecular mechanism of co-operation. I therefore focused on the interaction between *GATA1s* and a haploinsufficient cohesin gene mutation in haematopoietic cells with the aim of exploring, at a cellular and molecular level, how these two mutations together may be contributing to disease.

4

The impact of *Gata1s* and a cohesin gene mutation on haematopoiesis

Contents

4.1 Introduction

An appreciation of how the mutations seen in leukaemic cells co-operate to perturb normal haematopoiesis is key to understanding the molecular mechanisms underlying

leukaemogenesis. The work in the previous chapter showed that in ML-DS often only T21, a GATA1s mutation and a cohesin gene mutation are required for transformation to leukaemia, making it a good model in which to study mutational cooperativity (Nikolaev et al. 2013, Yoshida et al. 2013, Labuhn et al. 2019). I therefore set out to investigate the impact on haematopoiesis of a Gata1s mutation and a haploinsufficient cohesin gene mutation, both alone and in combination. Given the challenges of performing molecular experiments on primary human cells from low volume paediatric blood and bone marrow samples, I used a transgenic mouse model system.

4.1.1 Mouse haematopoiesis

The murine haematopoietic hierarchy has been extensively studied. It serves to generate a variety of specialised cell types during development and adult life. These mature haematopoietic cells have a limited life span and must be continually replenished. Broadly, this involves the self-renewal of pluripotent haematopoietic stem cells (HSCs) and their step-wise commitment to increasingly differentiated progenitor and precursor cell compartments. Different haematopoietic cell types are classically identified and separated on the basis of cell surface markers, and functionally characterised using *in vitro* and *in vivo* assays (Pronk et al. 2007, Orkin et al. 2008, Laurenti et al. 2018, Kiel et al. 2005).

4.1.1.1 Haematopoietic stem and progenitor populations

HSCs were first demonstrated *in vivo* in 1961 based on the ability of transplanted donor cells to rescue lethally irradiated recipient mice (Till and Mc, 1961). They were subsequently characterised as cells possessing both self-renewal and multipotency, purified using a variety of cell surface markers (Spangrude et al. 1988, Okada et al. 1992, Osawa et al. 1996). In the classical model of haematopoiesis (Figure 4.1a), HSCs are subdivided into two populations: long-term HSCs (LT-HSCs) and short-term HSCs (ST-HSCs) (Ogawa et al. 1991, Smith et al. 1991). LT-HSCs are rare and quiescent in steady state haematopoiesis. They are reactivated to enter the cell

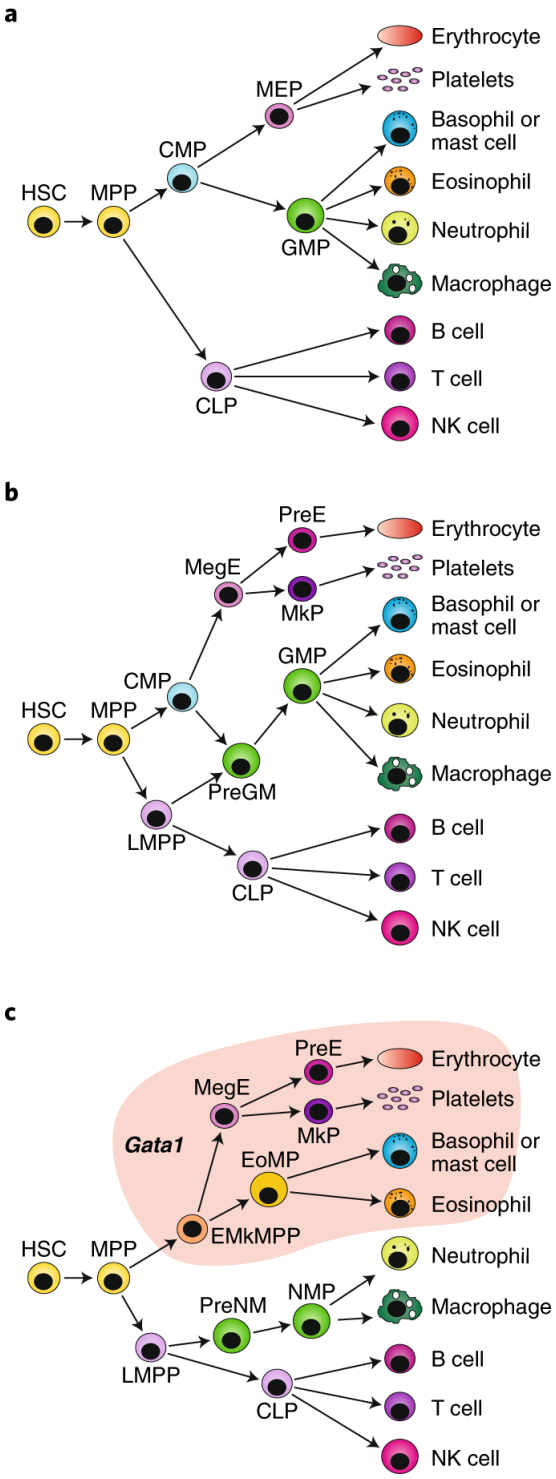


Figure 4.1: Models of mouse haematopoiesis. Schematic representation of CMP-CLP (a), LMPP-CMP (b) and EMkMPP- LMPP (c) models of mouse haematopoiesis. Shaded area indicates Gata1-expressing cells. MkP = megakaryocyte progenitor, MegE = pre-megakaryocyte-erythroid progenitor, preE = pre-colony-forming erythroid progenitor, EoMP = eosinophil-mast cell progenitor, NMP = neutrophil-monocyte progenitor, preNM = pre-neutrophil-monocyte progenitor, NK = natural killer. Adapted from Jacobsen and Nerlov, 2019 Nature Cell Biology with permission from Springer Nature.

cycle under stress conditions and possess the ability to reconstitute haematopoiesis in lethally irradiated recipient mice. ST-HSCs can reconstitute haematopoiesis in the short term and differentiate into multipotent progenitors (MPPs) with reduced self-renewal but increased proliferation and differentiation activity (Morrison et al. 1997). The first major lineage fate decision lies downstream of the MPPs, with differentiation either into common myeloid progenitors (CMPs) (which give rise to granulocyte-macrophage progenitors (GMPs) or megakaryocyte-erythroid progenitors (MEPs)) or common lymphoid progenitors (CLPs) and lymphoid-biased progeny (Kondo et al. 1997, Akashi et al. 2000).

This model has been revised following the identification of a subpopulation of HSCs - the lymphoid-primed multipotent progenitor population (LMPP) - with high Flt3 expression and combined lymphoid/granulocyte-macrophage potential but lacking erythroid-megakaryocyte potential (Adolfsson et al. 2005, Luc et al. 2008, Lai et al. 2006) (Figure 4.1b).

Single-cell lineage tracing experiments support a model in which lineage commitments develop in parallel within the MPP compartment (Lei et al. 2015), and work by the Trumpp and Passegue groups has allowed further resolution of the MPP populations into MPP1-4 based on immunophenotype, abundance in the bone marrow, differentiation potential and cell cycle status, with MPP1 sharing many of the characteristics of the ST-HSC and MPP2-4 being responsible for steady state haematopoiesis (Wilson et al. 2008, Pietras et al. 2015). *In vivo*, MPP2 was shown to represent a megakaryocyte-biased population, MPP3 a myeloid-biased population and MPP4 a lymphoid-primed population encompassing the LMPP. Following transplantation, HSCs were shown to first generate myeloid-biased MPP1 and MPP2, followed by lymphoid-primed MPP4. However, no MPP population was able to generate another MPP population (Wilson et al. 2008, Pietras et al. 2015).

To classify the myeloid progenitor (Lin-, Kit+, Sca1-) compartment, Pronk et al (2007) used additional markers to separate progenitors into MkP, GMP, PreMegE, PreGM, PreCFU-E and CFU-E (Pronk et al. 2007). However, single cell

transcriptional studies (Paul et al. 2015, Nestorowa et al. 2016) reveal heterogeneity within these populations.

Finally, within the HSC compartment, a platelet-biased HSC population has been identified based on von Willebrand Factor (vWF) expression (Sanjuan-Pla et al. 2013). The use of a Gata1-reporter has further resolved the granulocyte lineage, with identification of a Gata1-expressing fraction of the pre-GM population (termed erythroid-megakaryocyte primed multipotent progenitor- EMkMPP) which gives rise to megakaryocyte, erythroid, mast cell and eosinophil cells but not lymphoid or neutrophil-macrophage lineages (Drissen et al. 2016) (Figure 4.1c). Increased use of single cell technologies to profile these stem and progenitor cell compartments is likely to reveal further levels of heterogeneity (Jacobsen et al. 2019).

4.1.1.2 Megakaryocyte and erythroid differentiation

Downstream of the megakaryocyte-erythroid progenitor compartment, erythroid (Pronk et al. 2007, Koulnis et al. 2011, Palis 2014) and megakaryocyte (Nakorn et al. 2003, Ng et al. 2012, Kuhl et al. 2005) differentiation has been studied using immunophenotypic profiling, *in vitro* and *in vivo* assays. Megakaryopoiesis in particular has historically been challenging to investigate - largely due to the low frequency of megakaryocytes in bone marrow and other haematopoietic organs. The immunophenotyically defined Lin- CD16/32-low CD41+ Kit+ CD9+ megakaryocyte progenitor population first identified by Nakorn et al. has been shown to contain a subset of cells representing around 0.01% of total nucleated adult bone marrow cells. *In vitro* these cells gave rise mainly to colony forming unit megakaryocytes (CFU-Mk) and occasional burst forming unit megakaryocytes (BFU-Mk). On transplantation, they gave rise only to platelets and did not contribute to long term multilineage haematopoiesis (Nakorn et al. 2003, Kuhl et al. 2005). Interestingly, by tracking the progenitors produced by singly transplanted HSCs, Carrelha et al. demonstrated that a class of platelet-biased HSCs exist with a fate restricted to replenishment of megakaryocyte-related lineages prior to loss of multipotency. No such LT-HSC was identified with a fate restricted to any of the other blood lineages,

perhaps reflecting a need for ongoing replenishment of short-lived megakaryocyte progenitors (Carrelha et al. 2018).

A schematic overview of erythroid and megakaryocyte differentiation, together with expression levels of key cell surface markers is shown in Figure 4.2.

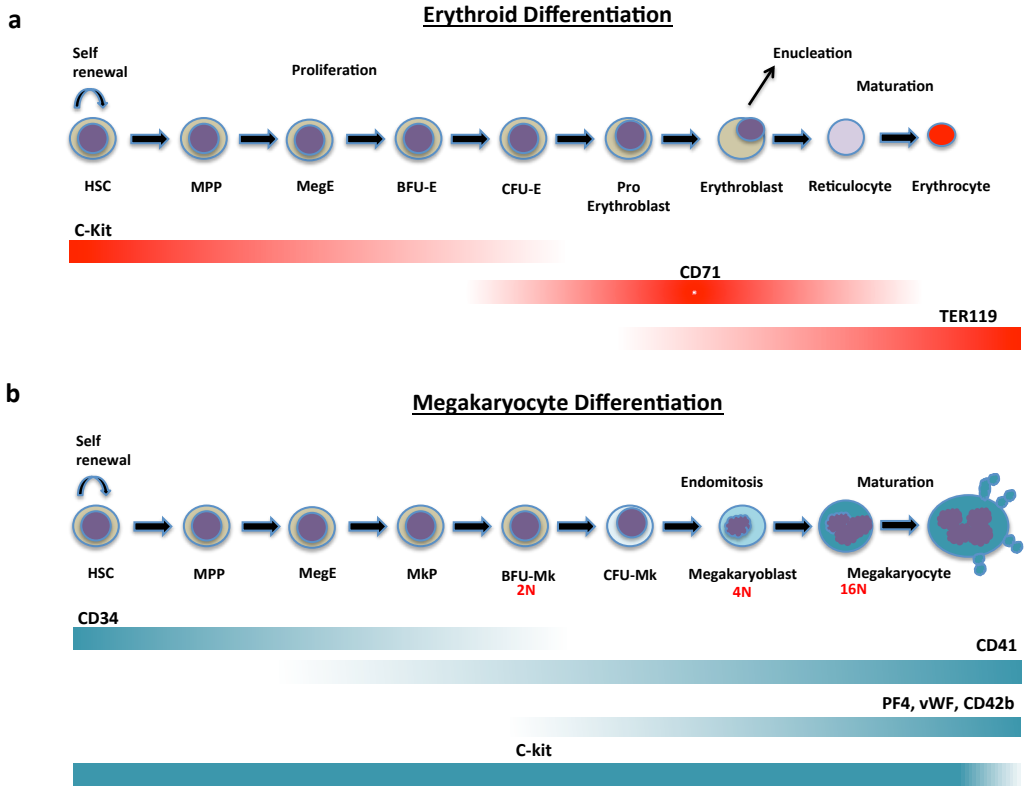


Figure 4.2: Erythroid and megakaryocyte differentiation. Simplified schematic of differentiation pathway of erythroid (a) and megakaryocyte (b) cells. Bars represent expression of cell surface markers. BFU = burst forming unit, CFU = colony forming unit, E = erythroid, Mk = megakaryocyte. N = ploidy.

4.1.2 Haematopoiesis in Gata1s mutant mice

The impact of exclusive production of Gata1s on haematopoiesis in mouse fetal liver was described by Li et al. (2005) as a predominantly embryonic phenotype, with transient reduction in TER119+ erythrocytes and colony forming erythroid (CFU-E) cells and increased numbers of hyperproliferative CD41+ megakaryocyte progenitors which had largely resolved by E14.5. This was associated with ~10% intraembryonic lethality due to fetal anaemia. Purified megakaryocyte progenitors from Gata1s mice were able to form proplatelets *in vitro*, but at a lower frequency than WT controls, suggesting a differentiation delay rather than complete differentiation block. By contrast, megakaryocyte progenitors from Gata1 null cells did not form proplatelets in culture (Li et al. 2005). These findings are supported by work done in the Vyas lab looking at haematopoietic differentiation in both ES and yolk sac cells from Gata1s and WT mice. Exclusive expression of Gata1s caused an arrest in erythroid differentiation and a differentiation delay in megakaryopoiesis, with accumulation of immature, Kit expressing, CD41-high megakaryocytic cells, possibly resulting from as few as three additional cell divisions in the mutant population (Juban et al. 2019 under review). To date, the bone marrow of adult Gata1s mice has not been studied in detail, however Alford et al. reported a mild macrocytic anaemia and thrombocytosis in adult Gata1s mutant mice, with splenomegaly developing in aged mice >15 months old (Alford et al. 2010). In the absence of T21 and cooperating mutations, none of these published Gata1s mouse models developed leukaemia.

4.1.3 Haematopoiesis in cohesin mutant mice

A number of studies have looked at the impact of cohesin haploinsufficiency on mouse haematopoiesis (Viny et al. 2015, Mazumdar et al. 2015, Mullenders et al. 2015, T. Wang et al. 2019). Perhaps unsurprisingly (given the importance of cohesin to dividing cells) haematopoietic specific homozygous loss of cohesin was lethal due to bone marrow failure (Viny et al. 2015, T. Wang et al. 2019). Heterozygous cohesin mutants were viable, fertile and did not develop leukaemia in the absence of cooperating mutations. They did develop a haematopoietic phenotype, characterised

in three of the four studies by a decrease in the LT-HSC compartment, increase in myeloid progenitor populations and increased replating efficiency and myeloid-skewed differentiation. Interestingly, contrasting the other studies, one group observed only a moderate decrease in myeloid populations in the cohesin mutant in steady state haematopoiesis, with a haematopoietic competitive disadvantage across all populations on transplantation. This was partially abrogated by concurrent *Dnmta3* haploinsufficiency, as seen in some adult AML patients, suggesting an important role for cooperativity between mutations in the disease phenotype (Wang et al. 2019). Differences to previous studies may, in part be explained by use of different cohesin targeting strategies. Key findings from the different studies are summarised in table 4.3, below.

Details of model	Immunophenotype	*In vitro* assays	*In vivo* assays	Reference
Conditional depletion of *Smc3* (Mx-Cre)	Homozygous knockout – lethal. Heterozygous knockdown → decrease in LT-HSC, increase in MPP. No change in myeloid progenitors.	Increased replating efficiency of heterozygous cells.	Increased serial transplantation in lethally irradiated recipients. Leukaemogenic in presence of Flt3-ITD.	Viny et al. 2013
Lentiviral transduction of cohesin mutants or shRNA knockdown in human CD34+ cells.	Myeloid skewing.	Increased replating efficiency.		Mazumdar et al. 2013
shRNA knockdown of cohesin components	Decrease in LT-HSC and ST-HSC. Trend towards increase in MPP2. No change in myeloid progenitors.	Increased replating efficiency of knockdown cells.	Aged mice developed myeloproliferative-like phenotype.	Mullenders et al. 2013
Conditional depletion of *Smc3* (Vav1-Cre and ERT2-Cre)	Homozygous knockout – lethal. Minimal effect of heterozygous knockdown.	No increased replating efficiency of heterozygous cells.	Decreased advantage on competitive transplantation.	Wang et al. 2019

Figure 4.3: Summary of cohesin knockdown studies in mice.

4.2 Aims

To date, the haematopoietic consequences of a *Gata1s* mutation together with haploinsufficiency of cohesin have not been described. In order to understand the mechanism by which these mutations might interact to perturb haematopoiesis, it was first important to robustly establish the resulting *in vivo* phenotype compared to wild type, *Gata1s* only and cohesin only mutant controls. I chose to focus on loss of function of the cohesin component Rad21 as it is: (a) one of the most commonly mutated cohesin components in ML-DS (Labuhn et al. 2019, Yoshida et al. 2013); (b) a core component of the cohesin ring and not subject to redundancy with a similar protein (as is potentially the case with SMC1A/SMC3 and STAG2/STAG1) and (c) the Rad21 mutations described in ML-DS are strongly suggestive of loss of function, making them more straightforward to model *in vivo*.

In overview, the aims for this section of work were:

1. To characterise the haematopoietic phenotype resulting from: (i) *Gata1s;* (ii) heterozygous knockdown of the core cohesin component *Rad21*; (iii) the combination of *Gata1s* and a *Rad21* mutation.

2. To assess the impact of any observed changes *in vitro*.

3. To determine whether, once a *Gata1s* mutation has been acquired, the addition of a cohesin gene mutation is sufficient for the development of leukaemia.

4.3 Results

4.3.1 Overview of transgenic mouse models

4.3.1.1 Gata1s mutant mice

The Gata1s mouse used in this study had previously been generated in the Vyas lab (Juban et al. 2019 under review). In this model, the *Gata1* gene was mutated by homologous recombination leading to production of a N-terminal truncated Gata1 protein tagged with a biotin tag, a TEV cleavage site and a FLAG tag (Figure 4.4a). In keeping with the *GATA1s* mutations seen in TAM and ML-DS, the targeting construct included most of the murine *Gata1* locus containing exon 2, intron 2 and a part of exon 3. Exclusive expression of *Gata1s* was confirmed by genotyping (Figure 4.4b). Exclusive production of Gata1s by Western blot has previously been demonstrated (Juban et al. 2019 under review). Gata1s mice were backcrossed onto a C57Bl/6 background. Mice were viable and fertile, although litter sizes were smaller than average for C57Bl/6 (Figure 4.4c) reflecting the previously reported ~10% embryonic lethality resulting from fetal anaemia (Li et al. 2005).

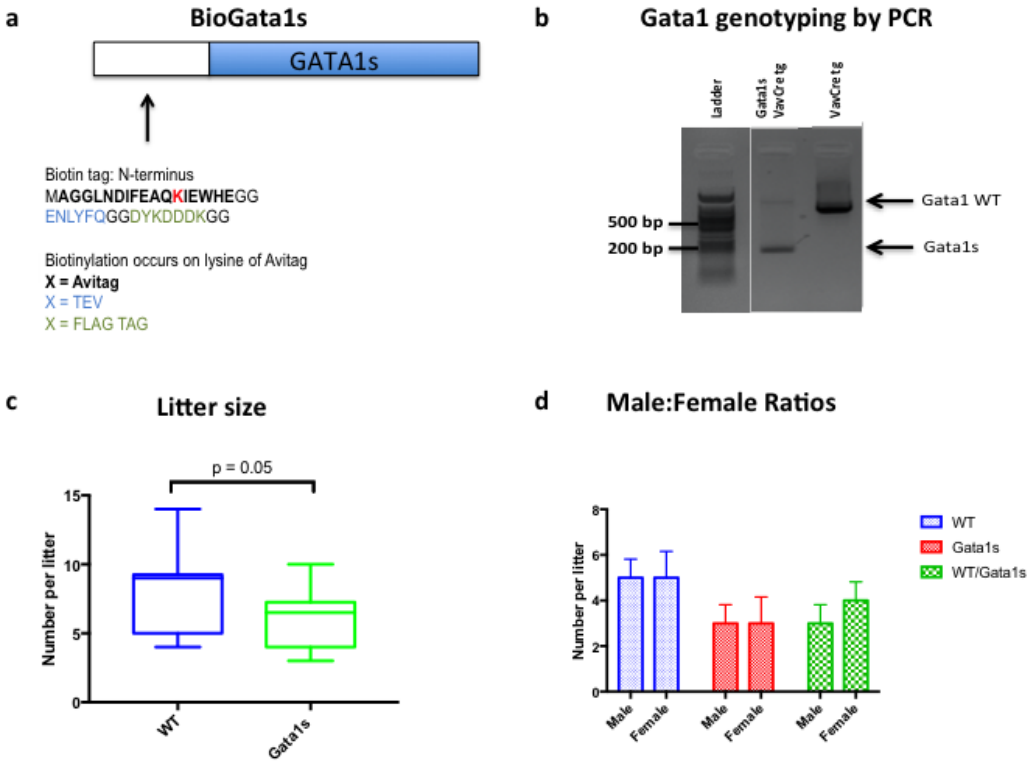


Figure 4.4: Gata1s mouse model. (a) Schematic of bioGata1s mouse model with N terminal truncation of the Gata1 protein and addition of a N terminal tag comprising Avitag, TEV cleavage site and FLAG tag. Biotinylation occurs on lysine of Avitag in presence of BirA. (b) PCR genotyping of Gata1 wildtype allele (802bp) and Gata1s allele (213bp). The genotype is shown for Gata1s VavCre tg and VavCre tg only. (c) Litter size of wild type and homozygous Gata1s mouse colonies (n=12 litters for each genotype). Box extends from 25th to 75th percentile. Line represents the median value. Whiskers extend to minimum and maximum values. Statistical analysis performed using Student's t-test. (d) Male:Female ratios in WT, homozygous Gata1s and WT/Gata1s litters (in which all male pups are Gata1s hemizygous and all female pups are Gata1s heterozygous.)

4.3.1.2 Rad21$^{fl/+}$VavCre tg mutant mice

To investigate the impact of heterozygous loss of Rad21 on the HSPC compartment, I bred Rad21$^{fl/+}$ mice (Figure 4.5a; Seitan et al. 2011) with VavCre mice (Boer et al. 2003). All mice were backcrossed onto a C57Bl/6 background using a speed congenic approach and the background was confirmed by SNP array (>98% C57Bl/6). Genomic excision of floxed *Rad21* exons 5 and 6 was demonstrated in bone marrow and spleen by PCR (Figure 4.5b) and depletion of Rad21 in Rad21$^{fl/+}$VavCre tg mutants confirmed by Western blot of mouse spleen and bone marrow cells (Figure 4.5c). As reported in other cohesin mutant mouse models (Mazumdar et al. 2015, Mullenders et al. 2015, Viny et al. 2015) no live Rad21$^{fl/fl}$VavCre tg pups were born, suggestive of early embryonic lethality. Rad21$^{fl/+}$VavCre tg mice were healthy to at least 52 weeks, fertile and morphologically normal. When bred as heterozygotes, litter sizes were normal.

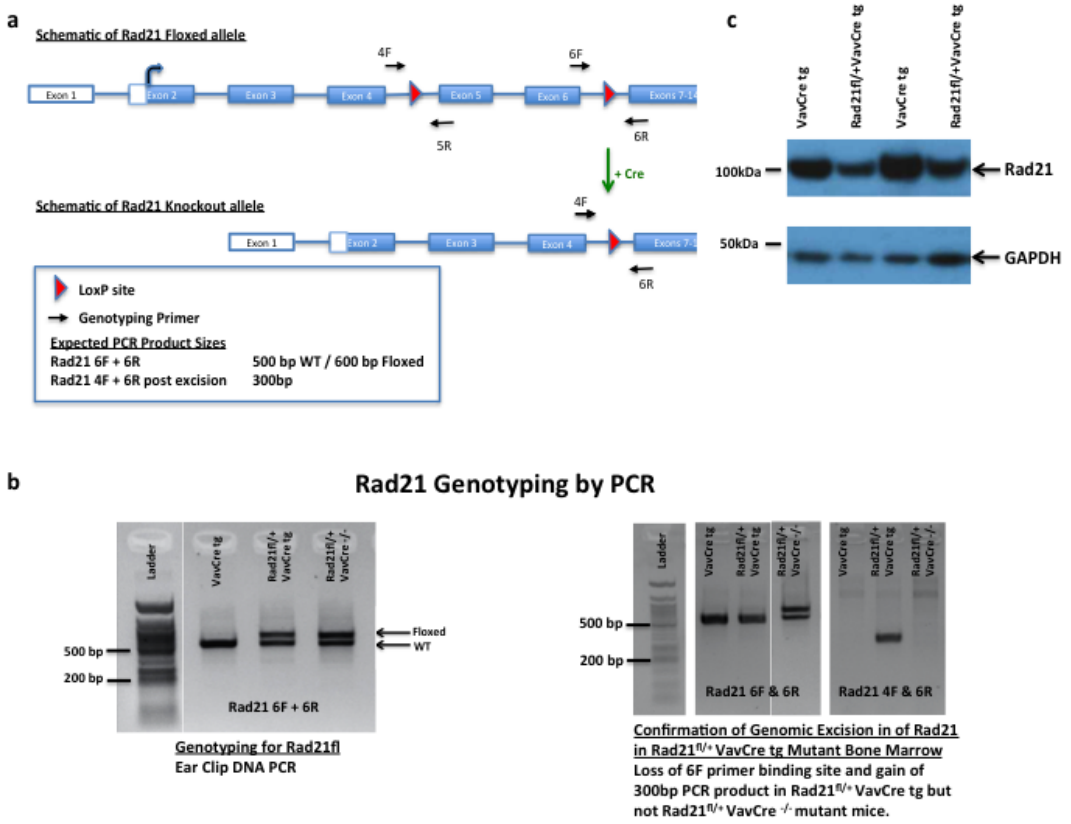


Figure 4.5: Rad21$^{fl/+}$VavCre tg mouse model (a) Schematic depiction of targeted Rad21fl allele. Black arrows indicate location of primers used for genotyping. (b) PCR genotyping of ear clip and bone marrow DNA for Rad21 wildtype allele (500bp), floxed allele (600bp) and recombination allele (300bp). The genotype is shown for VavCre tg, Rad21$^{fl/+}$VavCre tg and Rad21$^{fl/+}$VavCre$^{-/-}$. (c) Depletion of Rad21 in Rad21$^{fl/+}$VavCre tg mutants confirmed by Western blot of mouse spleen cells showing Rad21 band at 110kDa together with GAPDH loading control.

4.3.1.3 Gata1s together with heterozygous loss of Rad21: a combined mutant model

To create a combined Gata1s and Rad21$^{fl/+}$ mutant model, I bred Gata1s mice with with Rad21$^{fl/+}$VavCre tg mutants. Breeding strategy and summary of genotypes is outlined in Figure 4.6. Rad21 depletion and exclusive Gata1s expression was confirmed in the combined mutant (data not shown). As *Gata1* is on the X-chromosome, all reported analysis was performed on 6-12 week old male mice. Littermate controls were used where available. "WT" refers to VavCre tg or true wild type, "Rad" refers to Rad21$^{fl/+}$VavCre tg, "Gata" refers to Gata1s VavCre tg or Gata1s only and "Combined mutant" or "Comb" refers to the combined mutant: Rad21$^{fl/+}$VavCre tg Gata1s. No difference in described findings (immunophenotypic or molecular) was observed when comparing wild type vs VavCre tg only controls or when comparing Gata1s vs Gata1s VavCre tg controls so the data from these groups have been pooled for the purpose of analysis. Combined mutant mice were viable, fertile and indistinguishable in physical appearance from their littermates. Taking into account the reduction in litter size caused by Gata1s, combined mutant mice were born at a frequency in keeping with expected Mendelian ratios.

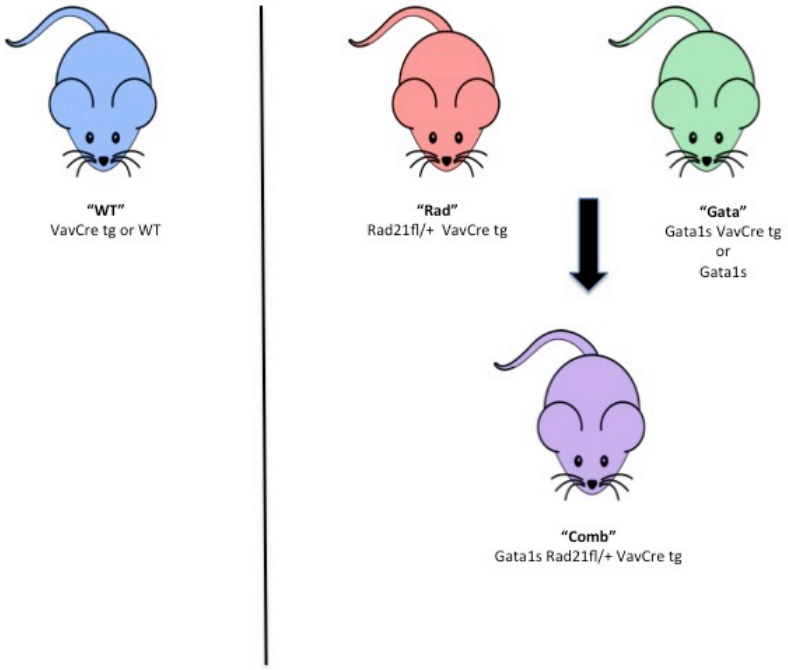


Figure 4.6: Mouse breeding strategy and genotypes analysed For the purposes of analysis, Gata1s Rad21 WT VavCre tg and Rad21 WT VavCre tg littermates were used as the "Gata" and "WT" controls where available. fl/+ represents a heterozygous floxed allele. Comb = combined mutant. WT = wild type.

4.3.2 Gata1s causes a macrocytic anaemia

Analysis of peripheral blood samples from 6-12 week old male mice revealed a subtle red cell phenotype in the combined mutant, with a statistically significant decrease in red cell count (p-value <0.0001, One-way ANOVA) and increase in mean cell volume (MCV) (p-value <0.0001, One-way ANOVA) when compared to the VavCre tg only wild type control (Figure 4.7).

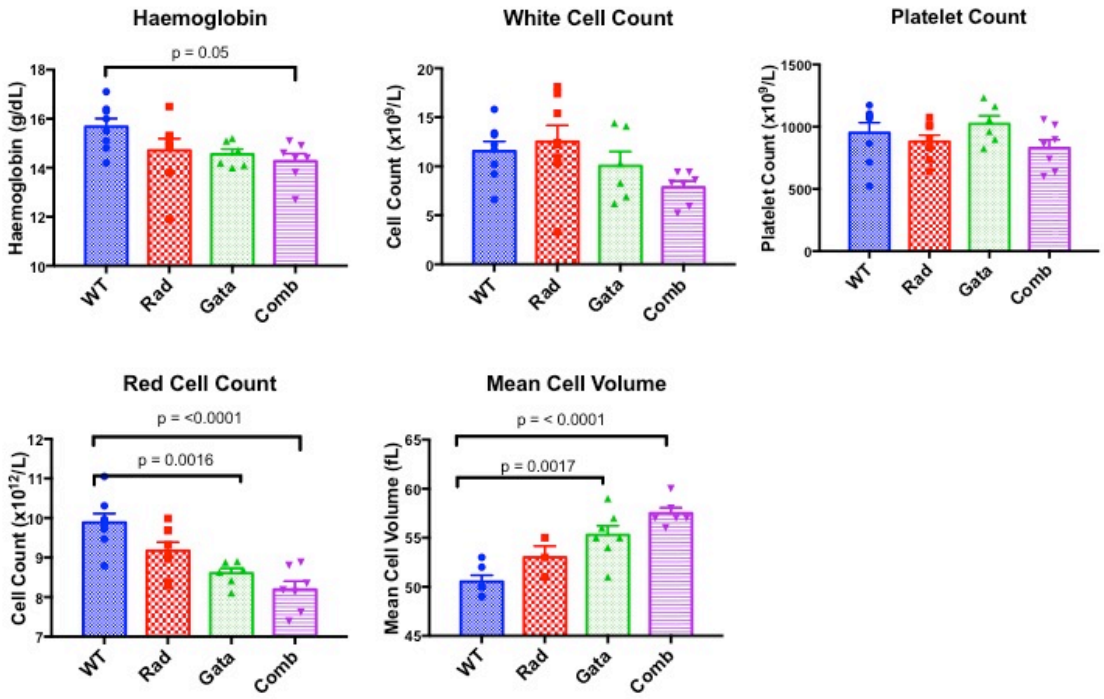


Figure 4.7: Summary of peripheral blood count parameters. Blood samples are from male mice aged 6-12 weeks. Each dot represents a separate biological sample. Graphs show Mean +/- SEM. Statistically significant changes are marked with corresponding p-values (One-way ANOVA).

The difference in red cell count and MCV between wild type and Gata1s-only mutant also reached statistical significance (p-values 0.0016 and 0.0017 respectively, One-way ANOVA) as previously reported in adult mice (Alford et al., 2010) suggesting a driving role for Gata1s in this erythropoietic phenotype. Although a downward trend was noted in white cell count in the combined mutant compared

to controls, this did not reach statistical significance. The platelet count was also unchanged from wildtype levels in both EDTA and citrated samples. However, blood film review revealed platelet clumping in the Gata1s mutant (Figure 4.8) (perhaps reflecting a degree of thrombocytosis that was missed by automated count), which was absent from the combined mutant and also WT and Rad21 only controls.

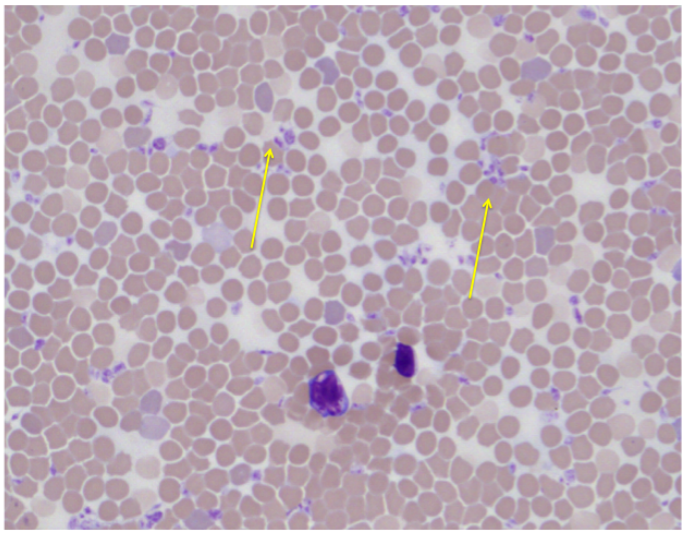


Figure 4.8: Platelet clumping in Gata1s mutant mice. Representative peripheral blood film from 8 week old Gata1s mutant mouse showing platelet clumps. Viewed at x40 magnification.

4.3.3 Gata1s plus cohesin haploinsufficiency co-operate to perturb haematopoiesis in the bone marrow haematopoietic stem and progenitor compartment

I next examined the bone marrow and spleen of WT, Rad, Gata and Combined mutant mice. Bone marrow cellularity, spleen cellularity and spleen size in Rad, Gata and Combined mutants were unchanged from wildtype (Figure 4.9).

Immunophenotypic analysis of bone marrow stem and early progenitor populations was performed. A representative gating strategy is shown in Figure 4.10.

In keeping with previously published mouse models of cohesin deficiency (Viny et al. 2015, Mazumdar et al. 2015, Mullenders et al. 2015), a significant decrease in the frequency of HSCs (Lin- Sca1+ Kit+ Flt3- CD150+ CD48-, P=0.01 One-way

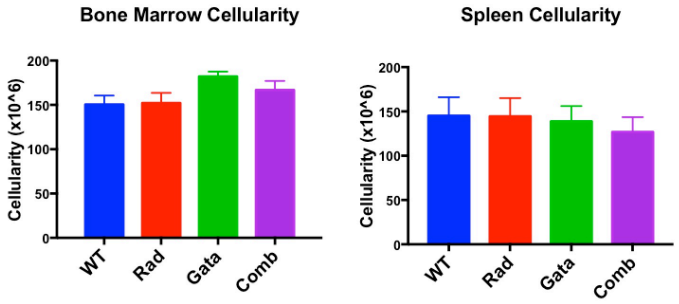


Figure 4.9: Bone marrow and Spleen cellularity. Samples are from male mice aged 6-12 weeks. N>6 for all four genotypes. Graphs show Mean +/- SEM. No statistically significant differences were noted between groups by One-Way ANOVA.

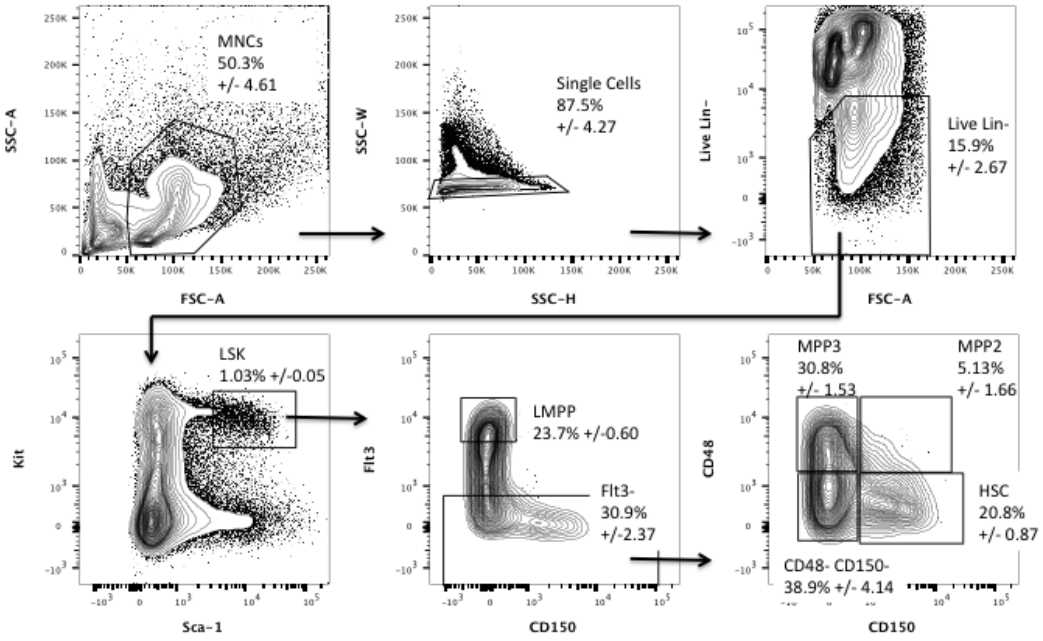


Figure 4.10: Immunophenotypic analysis of murine haematopoietic stem cell and early progenitor populations: gating strategy. Gates labelled with name of population and mean % of preceding population +/-standard deviation in WT samples. MNCs=mononuclear cells, Lin = lineage, LSK = Lin-Sca1+c-Kit+ cells, LMPP = lymphoid-primed multipotent progenitors, HSC=haematopoietic stem cell, MPP=multipotent progenitors.

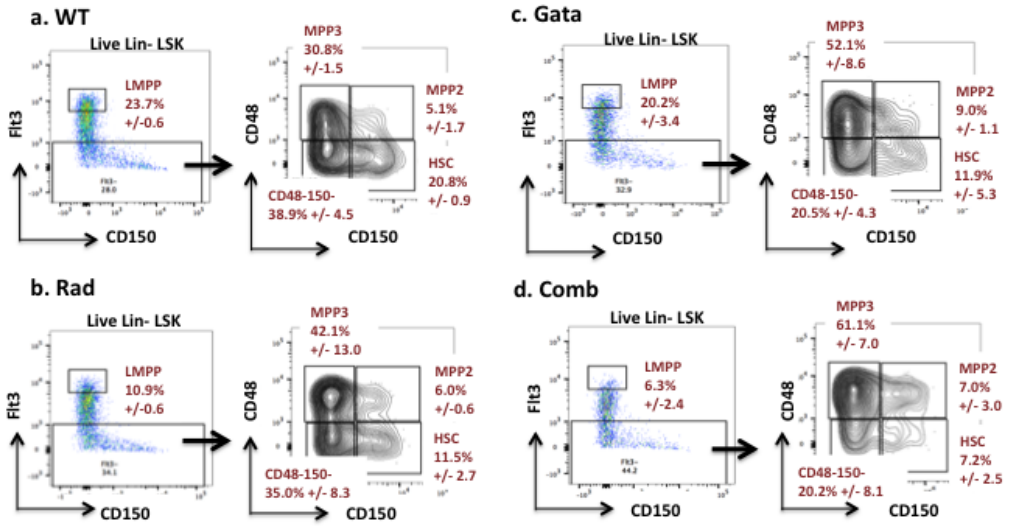


Figure 4.11: Immunophenotypic profile of stem and progenitor populations. Representative FACS profiles of haematopoietic stem and progenitor populations from the bone marrow of wild type (WT) (a), cohesin mutant (Rad) (b), Gata1s (Gata) and combined Gata1s cohesin mutant mice (Comb) (d). Gates labelled with the name of population and mean % of preceding population +/-standard deviation in WT samples. Lin = lineage, LSK = Lin-Sca1+Kit+, LMPP=lymphoid-primed multipotent progenitors, MPP=multipotent progenitor population, HSC=haematopoietic stem cell.

ANOVA) was noted when comparing WT and Rad mutant mice (Figure 4.9 and Figure 4.10). In the combined mutant, the decrease in HSC was more profound (P=0.002, One-Way ANOVA). Interestingly this compartment was also decreased in the Gata1s only mutant. The size of the Lin- Sca1+ Kit+ Flt3- CD150- CD48- compartment was also reduced in the combined mutant compared to controls (P=0.01, One-way ANOVA). There was a trend towards expansion of the myeloid progenitor populations in combined mutant mice: MPP3 (Lin- Sca1+ Kit+ Flt3- CD150- CD48+), shown to have a predominantly GM potential in steady state haematopoiesis (Pietras et al. 2015, Wilson et al. 2008), and MPP2 (Lin- Sca1+ Kit+ Flt3- CD150+ CD48+) which has a predominantly megakaryocyte/erythroid

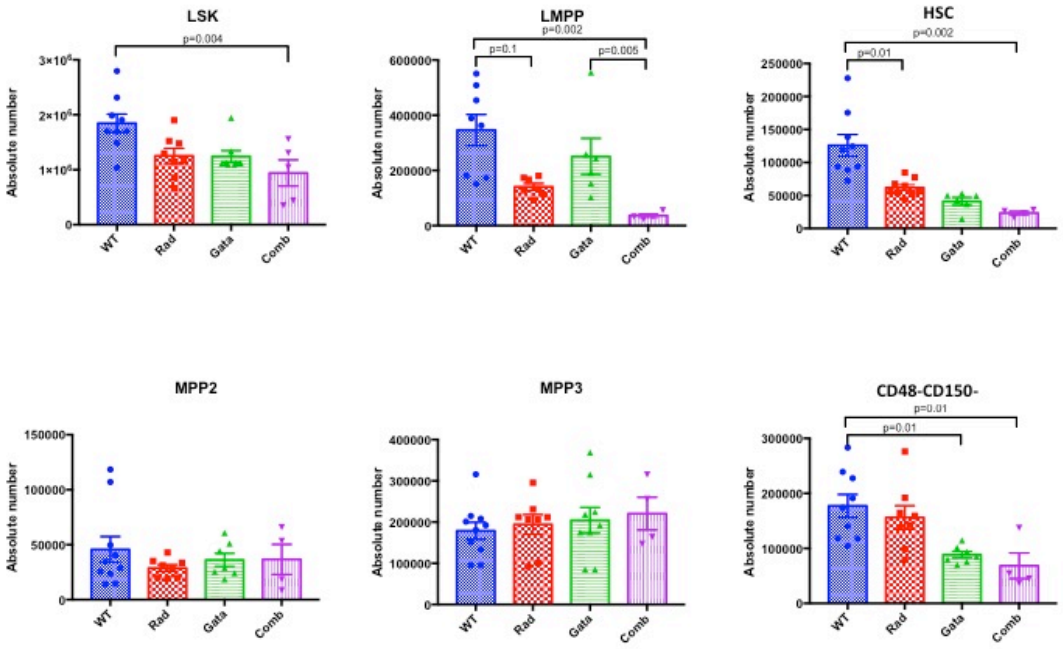


Figure 4.12: Decreased lymphoid-primed multipotent progenitor and HSC compartments in combined mutant mice. Absolute size of the indicated stem and progenitor populations in the bone marrow cells of 6-12 week old male mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

potential (Pietras et al. 2015) (Figure 4.11) however this did not reach statistical significance when absolute numbers in each compartment were calculated (Figure 4.12). The lymphoid-primed multipotent progenitor population (LMPP: Lin- Sca1+ Kit+ Flt3-hi) was significantly decreased in both Rad only and combined mutant mice (P=0.002, One-Way ANOVA).

4.3.4 The effect of Gata1s plus cohesin haploinsufficiency on the myeloid progenitor compartments

I investigated the frequency of myeloid progenitors in the bone marrow of WT, Rad, Gata and Comb mutant mice using the Bryder panel (Pronk et al. 2007). A representative gating strategy is shown in Figure 4.13.

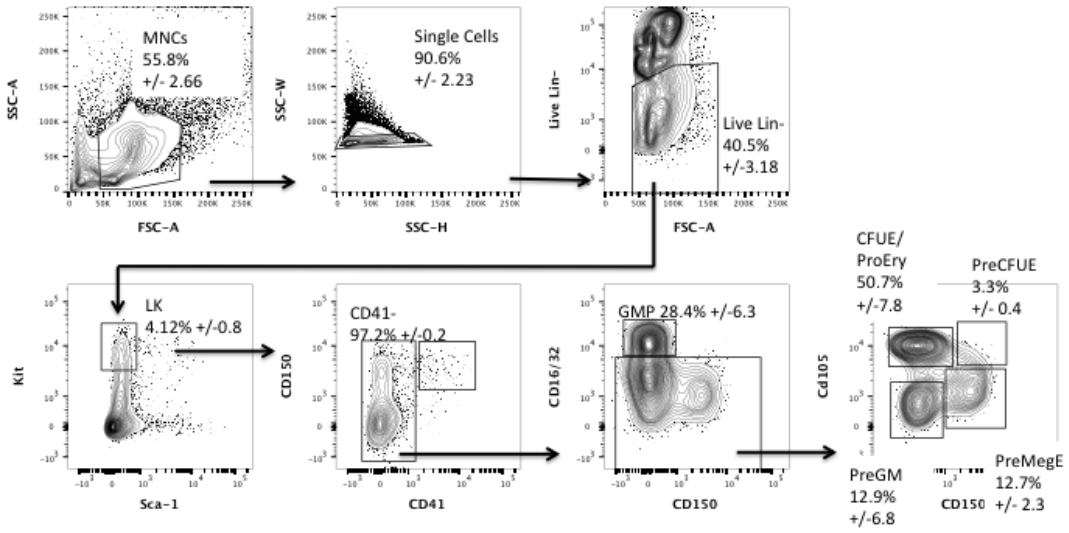


Figure 4.13: Immunophenotypic analysis of murine myeloid progenitors (Bryder panel): gating strategy. Gates labelled with name of population and mean % of preceding population +/-standard deviation in WT samples. MNCs=mononuclear cells, Lin = lineage, GMP = granulocyte-monocyte progenitor, CFU-E = colony forming unit erythroid cells, pre-MegE = pre-megakaryocyte erythroid progenitor, proEry = pro-erythroblasts.

No statistically significant differences were noted in the size of the granulocyte monocyte progenitor (GMP), Pre-GM, Pre-CFU-E, CFU-E&Pro-Ery or pre-megakaryocyte erythroid (PreMegE) progenitor populations in Rad, Gata or Comb mutant mice compared to WT (Figure 4.14), although a trend towards expansion of the PreMegE and reduction in size of CFU-E&Pro-Ery in Gata and Comb mutants compared to WT and Rad mice was noted.

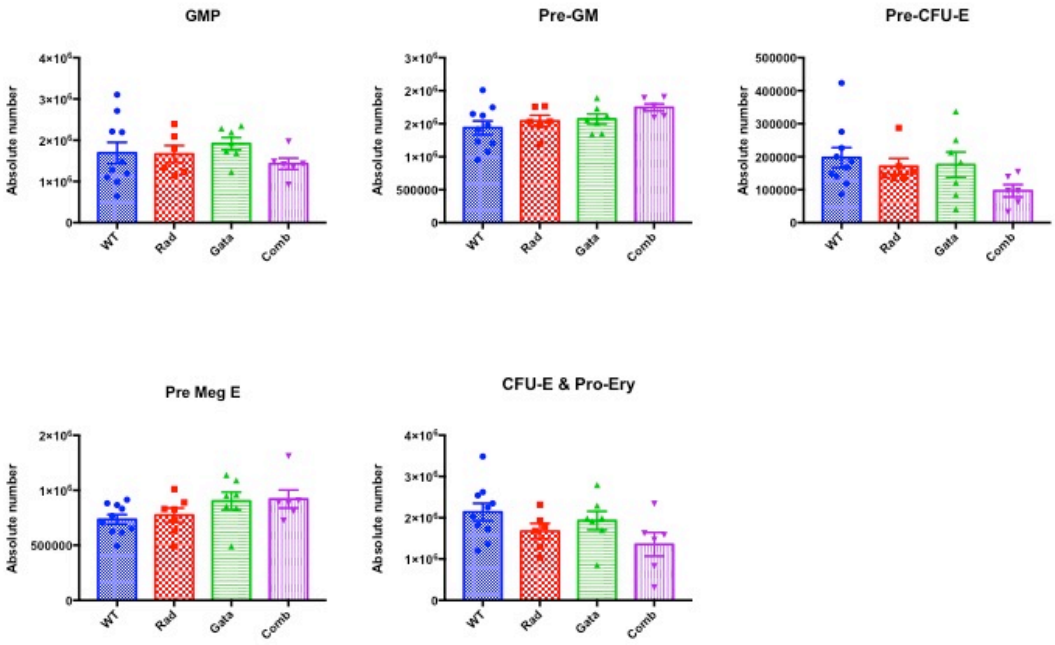


Figure 4.14: Absolute size of myeloid progenitor compartments is unchanged in Gata1s, cohesin mutant and combined mutant mice compared to wild type. Absolute size of the indicated stem and progenitor populations in the bone marrow cells of 6-12 week old male mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. No statistically significant differences were seen (One-way ANOVA).

4.3.5 Gata1s plus cohesin haploinsufficiency may impair erythropoiesis in the bone marrow

To explore the erythroid phenotype in more detail, the erythroid marker TER119 was added to the Bryder panel, allowing immunophenotypic separation of CFU-E (colony forming unit erythroid cells) and Pro-Ery (pro-erythroblasts)(Pronk et al. 2007). A trend towards decrease in size of the Pre-Ery and CFU-E compartments was seen in the combined mutant compared to wild type control, although this did not reach statistical significance by One-Way ANOVA (Figure 4.15 and Figure 4.16). A statistically significant decrease was seen in the TER119-hi cell population in Comb mutant mice (and to a lesser extent Gata1s only mutant mice) compared to WT controls (P=0.0095, One-Way ANOVA) suggesting a possible effect of Gata1s and Cohesin gene mutations on late erythropoiesis.

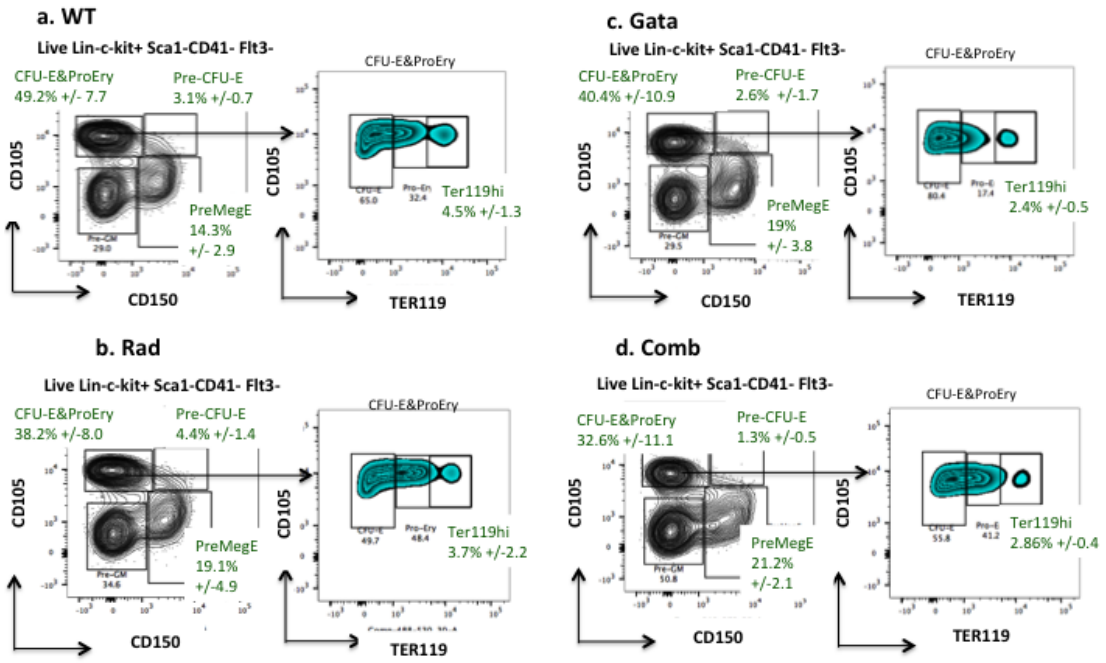


Figure 4.15: Immunophenotypic profile of red cell precursors. Representative FACS profiles of red cell precursor populations from the bone marrow of wild type (WT) (a), cohesin mutant (Rad) (b), Gata1s (Gata) (c), and combined Gata1s/cohesin mutant mice (Comb) (d). Gates labelled with the name of population and mean % of preceding population +/-standard deviation. Lin = lineage. CFU-E = colony forming unit erythroid cells, pre-MegE = pre-megakaryocyte erythroid progenitor, proEry = pro-erythroblasts.

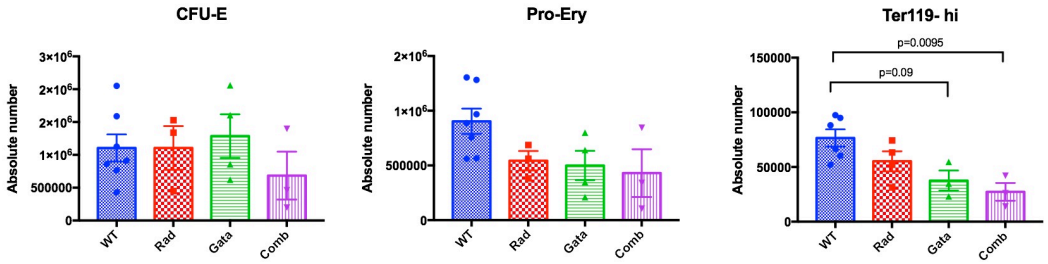


Figure 4.16: Erythroid progenitor compartments in wild type, Gata1s, cohesin mutant and combined mutant mice. Absolute size of the indicated progenitor populations in the bone marrow cells of 6-12 week old male mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. All statistically significant differences are marked (One-way ANOVA, p-value shown).

4.3.6 Gata1s and cohesin haploinsufficiency cause expansion of the megakaryocyte progenitor compartment

The megakaryocyte progenitor population (MkP) was defined immunophenotypically as Lin-Sca1-CD16&32-CD41+Kit+CD9+ (Nakorn et al. 2003) (Figure 4.17). A significant increase in the size of this compartment was observed in Combined mutant mice, compared to WT (P=0.0001, One-Way ANOVA), Rad and Gata1s only (P=0.0015, One-Way ANOVA) controls (Figure 4.19). Despite an increased number of cells in the CD9+ MkP gate of the combined mutant, both by percentage of preceding population and absolute count, this was accompanied by a reduction in CD9 surface expression and characteristic FACS plot appearance (Figure 4.18).

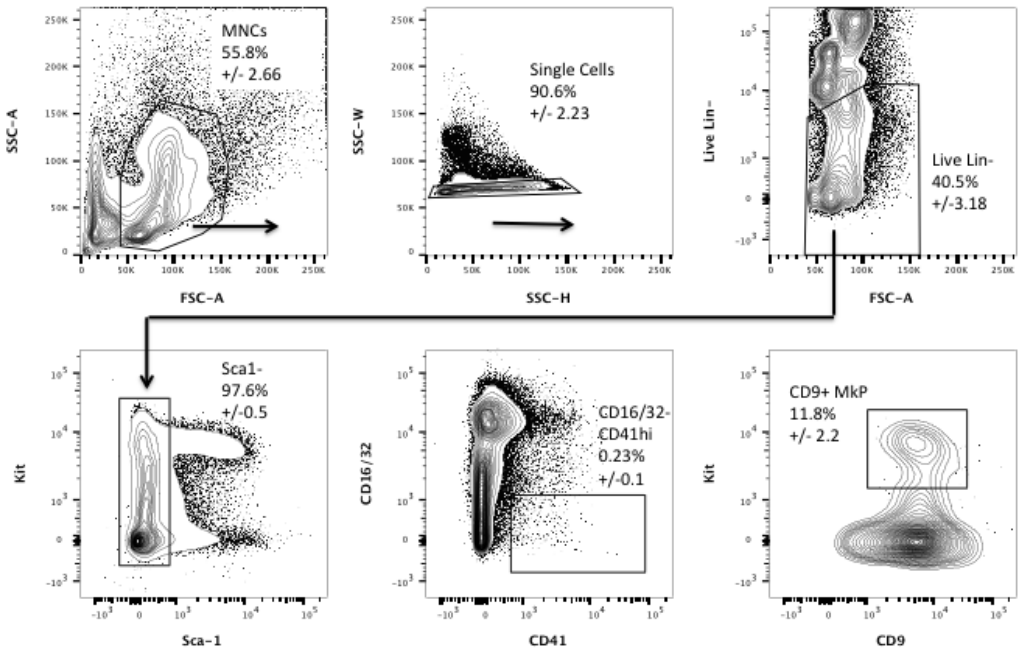


Figure 4.17: Immunophenotypic analysis of CD9+ murine megakaryocyte progenitor population: gating strategy. Gates labelled with name of population, mean% of preceding population +/-standard deviation in WT. MNCs=mononuclear cells, Lin = lineage, MkP = megakaryocyte progenitor population.

Given that ML-DS is morphologically a megakaryocytic leukaemia, distinguishable from TAM by a differentiation block and thrombocytopenia, this novel MkP

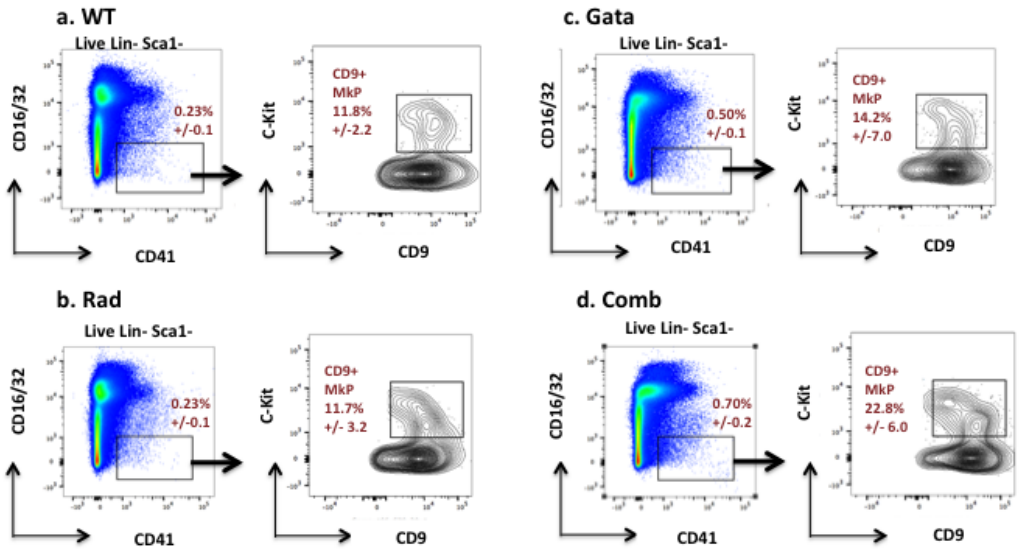


Figure 4.18: Immunophenotypic profile of megakaryocyte progenitor population (CD9+ MkP). Representative FACS profiles of the CD9+ megakaryocyte progenitor population from the bone marrow of wild type (WT) (a), cohesin mutant (Rad) (b), Gata1s (Gata) (c), and combined Gata1s cohesin mutant mice (Comb) (d). Gates labelled with the name of population and mean % of upstream population +/- standard deviation.

phenotype is interesting and potentially of clinical relevance.

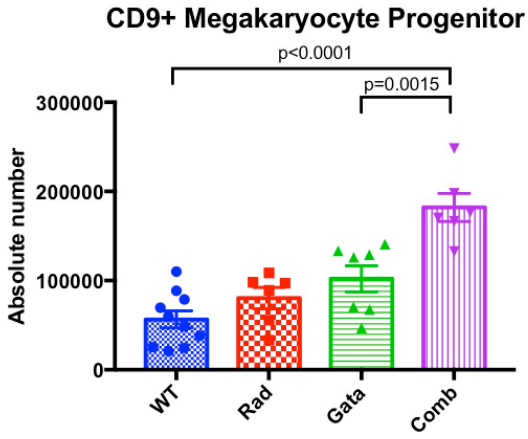


Figure 4.19: Gata1s plus a cohesin gene mutation results in an increase in the CD9+ megakaryocyte progenitor compartment in mouse bone marrow. Absolute size of the indicated progenitor populations in the bone marrow cells of 6-12 week old male mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

Prior to embarking on further experiments, I wanted to ensure that this was a robust phenotype, independent of background strain and the presence of an N-terminal biotag in our Gata1s and Comb mutant mouse models. I therefore repeated the analysis using Gata1s mice developed by the Orkin lab (Li et al. 2005). These mice exclusively express the Gata1s isoform but have no associated tag on the truncated N-terminus. Orkin-Gata1s mice were bred with Rad21fl/+ VavCre tg mutants to produce WT, Gata1s and Comb mutant mice on the same background and bone marrow analysis was performed as previously described in 6-12 week old male littermates. As shown in Figure 4.20 and Figure 4.21, statistically significant expansion of the CD9+ MkP in the combined mutant was observed.

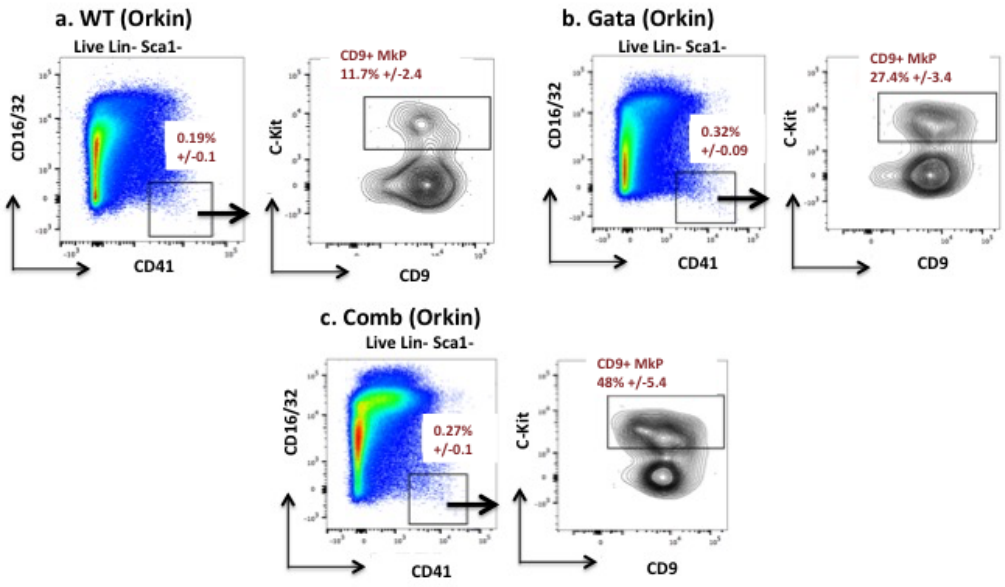


Figure 4.20: Immunophenotypic profile of megakaryocyte progenitor population (CD9+ MkP) using alternative Gata1s mouse model. Representative FACS profiles of the CD9+ megakaryocyte progenitor population from the bone marrow of wild type (WT) (a), Gata1s (Orkin) (b), and combined Gata1s (Orkin) cohesin mutant mice (Comb) (c). All mice were on a mixed background and littermate controls. Gates labelled with the name of population and mean % of upstream population.

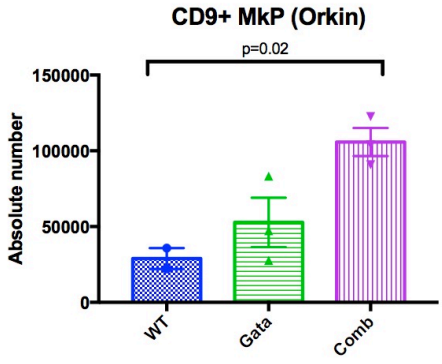


Figure 4.21: Increase in the CD9+ megakaryocyte progenitor compartment in combined mutant mice is also seen using an alternative Gata1s mouse model. Absolute size of the indicated progenitor populations in the bone marrow cells of 6-12 week old male mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

4.3.7 Gata1s and cohesin haploinsufficient mice have increased megakaryocytes in the bone marrow

To determine whether the observed expansion in the MkP compartment in combined mutant mice was accompanied by an increase in megakaryocytes in the bone marrow, I examined H+E stained bone marrow trephines from the femurs of 6-12 week old WT, Rad, Gata and Comb mice (n=3 for each genotype). A significant increase in the number of megakaryocytes was noted in the Combined mutant compared to WT, Rad and Gata controls (Figure 4.22 and Figure 4.23). Morphologically, Gata1s-only megakaryocytes were typically larger than wild type with multi-lobulated nuclei. By contrast, in the Combined mutant, megakaryocytes tended to be smaller and hypo-lobated (Figure 4.23).

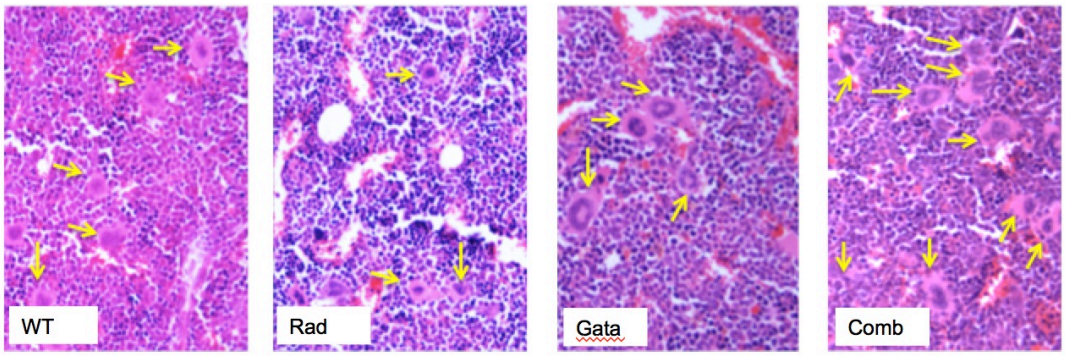


Figure 4.22: Representative bone marrow trephine sections showing increased numbers of megakaryocytes in Combined mutant mice compared to controls. Megakaryocytes are marked with yellow arrows. Staining with H+E. x40 magnification.

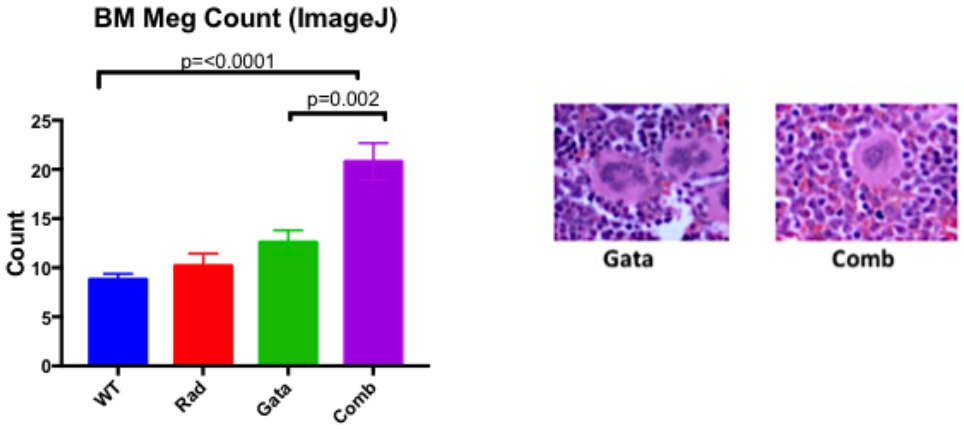


Figure 4.23: Number of megakaryocytes is increased in bone marrow of combined Gata1s/cohesin mutant mice. Bone marrow sections (n=3 for each genotype and 20 fields selected at random). Megakaryocytes were counted using ImageJ software. Graph represents count per field. Bars show standard error of the mean. Statistically significant differences and p values are shown (One-way ANOVA). On right, difference in morphology between megakaryocyte from Gata1s only and Combined mutant mice.

4.3.8 Gata1s and heterozygous loss of Rad21 cooperate to alter megakaryocyte progenitor proliferation and differentiation

Sorted CD9+ MkP cells were next cultured in medium containing TPO, EPO, IL-3, IL-11, SCF mFlt3-Ligand and GM-CSF (Kuhl et al. 2005). By day 3, WT CD9+ MkP progenitors (Lin-Sca1-Kit+CD16/32-CD41+CD9+) had differentiated into large cells with pro-platelets visible on the cell surface. WT GMP cells (Lin- Sca1+ Kit+ CD41- CD150+), sorted as a control population and cultured under identical conditions, did not display this phenotype (Figure 4.24). Cultured WT CD9+ MkP cells expressed high levels of the megakaryocyte marker CD41 and low levels of the erythroid marker TER119 and the stem marker Kit (Figure 4.25).

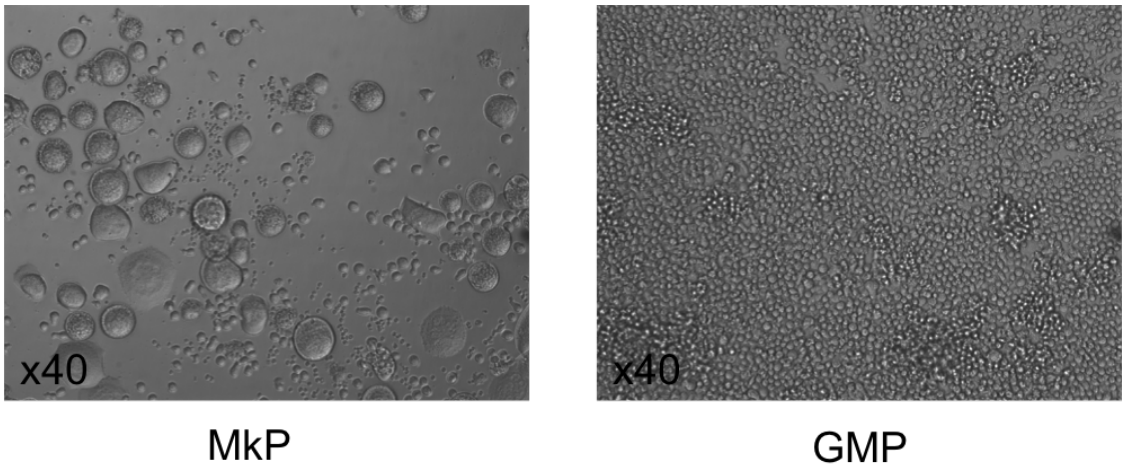


Figure 4.24: Sorted MKP and GMP cells photographed on D3 of liquid culture 3000 sorted cells were cultured in medium containing TPO, EPO, IL-3, IL-11, SCF mFlt3-Ligand and GM-CSF at 37°C for 3 days. MkP (left) differentiated into large cells with proplatelet extensions. GMP cells cultured as a negative control (right) remained small and no proplatelets were visible. Photographed at x40 magnification.

When equal numbers of CD9+ MkP cells were sorted and cultured from WT, Rad, Gata and Comb mutant mice under the above conditions, cells from both the Gata1 only and combined mutant mice displayed altered proliferation compared to controls (Figure 4.25 and Figure 4.26) with an increase in total cell count and absolute number of CD41+ cells most marked in the combined mutant (P=0.001 compared to WT, One-Way ANOVA).

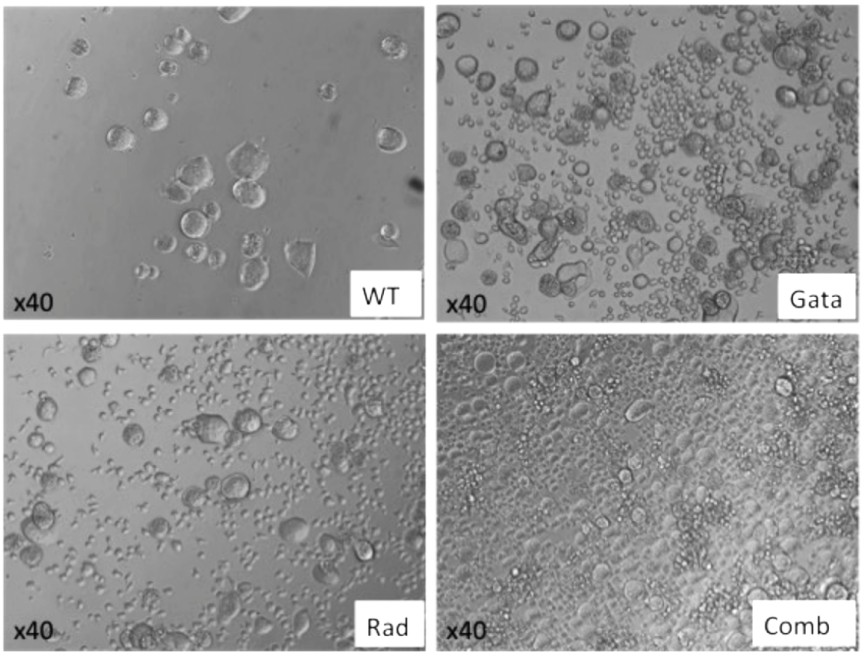


Figure 4.25: Sorted MKP cells from WT, Rad, Gata and Comb mutant mice 3000 sorted cells were cultured in medium containing TPO, EPO, IL-3, IL-11, SCF mFlt3-Ligand and GM-CSF at 37°C for 3 days. Photographed at x40 magnification.

Differentiation of these cells was also affected, with a proportional decrease in the megakaryocyte maturation markers CD41 (P=0.01 Comb v WT One-Way ANOVA) and CD42b and increased expression of Kit (P=0.05 Comb v WT One-Way ANOVA) and TER119 (P=0.003 Comb v WT One-Way ANOVA) in the combined mutant, perhaps suggestive of a less mature population retaining a degree of bi-potentiality (Figure 4.27). Representative FACS plots are shown in Figure 4.28.

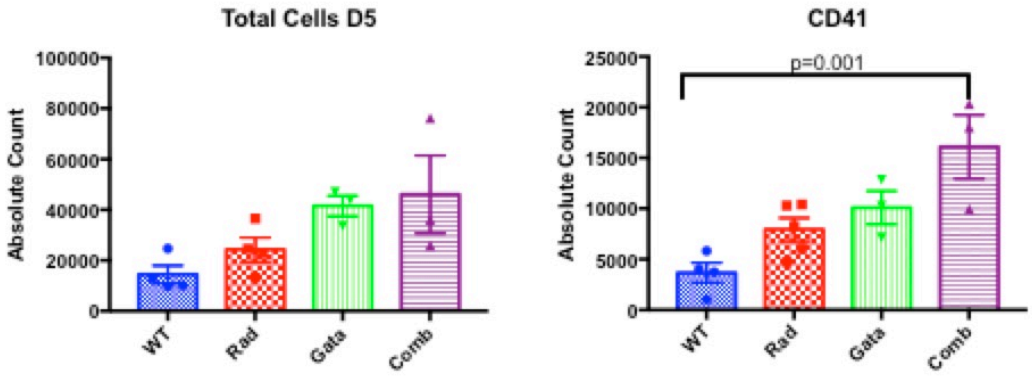


Figure 4.26: Gata1s and heterozygous loss of Rad21 cooperate to alter megakaryocyte progenitor proliferation *in vitro*. For each genotype, 3000 CD9+ MkP cells were sorted and cultured at 37°C. Absolute number of cells and number of CD41+ cells recorded on D5. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

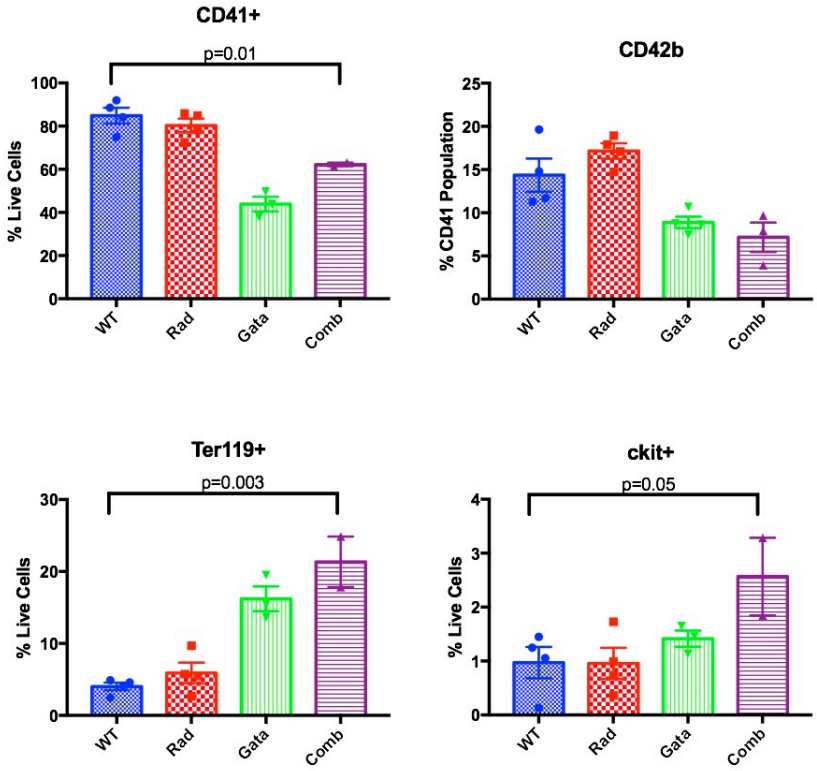


Figure 4.27: Gata1s and heterozygous loss of Rad21 cooperate to alter megakaryocyte progenitor differentiation *in vitro* For each genotype, 3000 CD9+ MkP cells were sorted and cultured in medium containing TPO, EPO, IL-3, IL-11, SCF mFlt3-Ligand and GM-CSF at 37oC. Graphs represent % of live cells with the exception of CD42b which is shown as % of parent CD41+ population. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

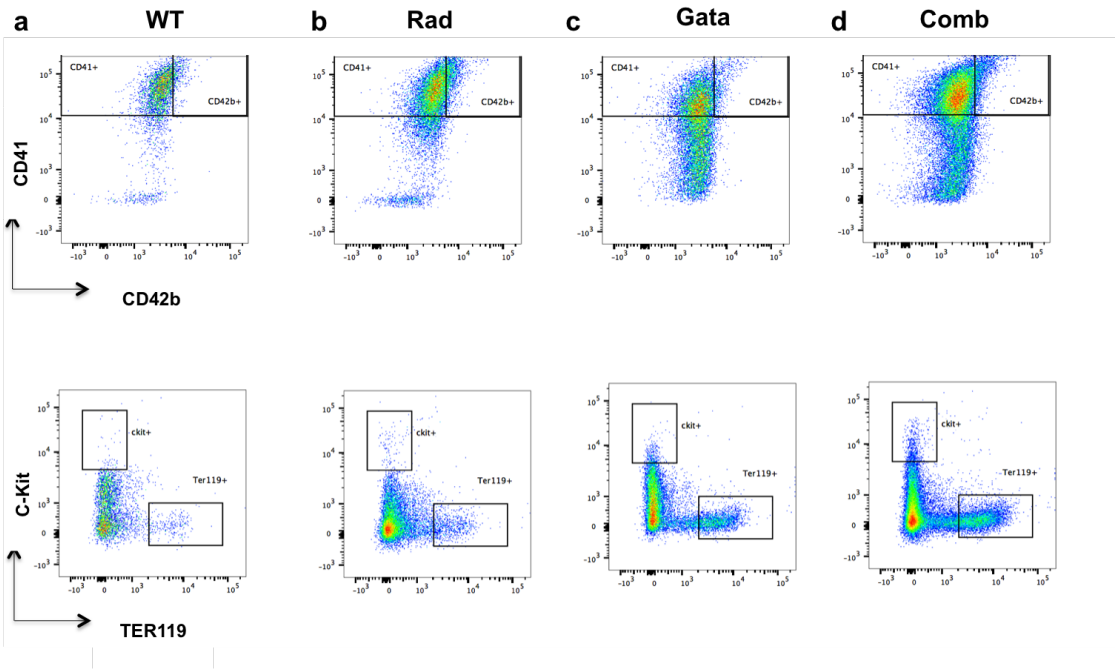


Figure 4.28: Immunophenotypic profile of sorted CD9+ MkP population after 5 days in liquid culture. Representative FACS profiles of MkP sorted from wild type (WT) (a), cohesin mutant (Rad) (b), Gata1s (Gata) (c), and combined Gata1s cohesin mutant mice (Comb) (d).

Ploidy was assessed in the CD41/CD42b double positive population. Analysis was limited due to low absolute numbers from all genotypes. However, in keeping with the qualitative morphological changes noted on assessment of bone marrow trephine, ploidy was increased in the Gata1s only mutant compared to WT (P=0.02 One-Way ANOVA), but not the combined mutant. There was also a possible trend towards decreased ploidy in the cohesin only mutant, however this did not reach statistical significance (Figures 4.28 and 4.29).

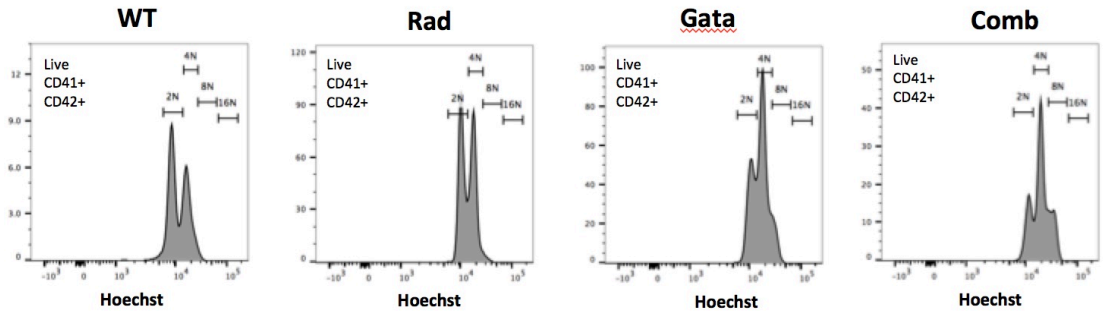


Figure 4.29: Ploidy analysis in CD41/CD42b cells. Representative FACS profiles of MkP sorted from wild type (WT) (a), cohesin mutant (Rad) (b), Gata1s (Gata) and combined Gata1s cohesin mutant mice (Comb) (d). Live, CD41/42b positive population is shown. Gates labelled with name of population. Y-axis represents count.

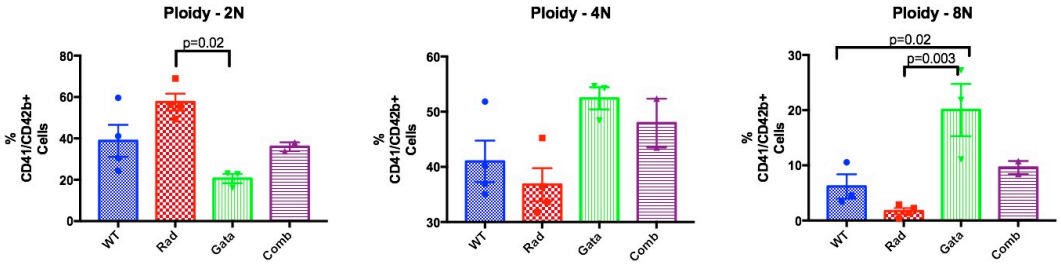


Figure 4.30: Increased ploidy in Gata1 mutant cells, but not the combined Gata1s/cohesin haploinsufficient mutant. Graphs represent % of live CD41/CD42b+ cells for each genotype after 5 days of in vitro culture. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

]

4.4 Discussion

Gata1s and haploinsufficient cohesin gene mutations co-exist in patients with ML-DS. How they cooperate to perturb haematopoiesis and contribute to myeloid leukaemogenesis has not been previously studied. The data presented in this chapter demonstrate that in a transgenic mouse model, Gata1s and haploinsufficiency of the cohesin component Rad21 together result in a novel haematopoietic phenotype with marked expansion of the bone marrow megakaryocyte progenitor compartment and evidence of abnormal proliferation and differentiation of mutant MkP cells.

My first aim was to characterise the impact of Gata1s and a haploinsufficient Rad21 mutation alone and in combination on murine haematopoiesis.

In keeping with what is known about Gata1s (Li et al. 2005, Kuhl et al. 2005, Juban et al. 2019 under review), Gata1s mutant mice demonstrated a subtle red cell defect with a mild macrocytic anaemia and decrease in the pro-ery and TER119-high erythroid populations in the bone marrow. There was a skew towards megakaryocyte proliferation with platelet clumping on blood film (perhaps reflective of a degree of thrombocytosis), moderate expansion of the PreMegE and MkP compartments and increase in hyperlobated megakaryocytes in the marrow. Cultured Gata1s MkP cells were more proliferative than wild type and showed reduced expression of mature megakaryocyte markers. Interestingly, despite an apparent impairment in terminal erythropoiesis in vivo, sorted Gata1s MkP cells demonstrated expression of the erythroid marker TER119 after 5 days in culture. This could represent an increase in erythroid differentiation and/or increased proliferation of TER119 expressing cells but has not been explored in detail here. Equally unexpectedly (given that *Gata1* expression would not be predicted in these compartments), a significant decrease in the LT-HSC and ST-HSC was noted in Gata1s mutants compared to *Gata1fl* WT controls.

Consistent with the work of others (Viny et al. 2015; Mazumdar et al. 2015; Mullenders et al. 2015; Wang et al. 2019), the haematopoietic effects of Rad21 haploinsufficiency alone were relatively subtle. A significant decrease in the size of both the LT-HSC and LMPP compartments were noted in cohesin mutant

mice, with possible expansion of the MPP compartments, however these did not reach significance in this study. Although the absolute number of megakaryocyte progenitors in cohesin mutant bone marrow did not significantly change from WT, Rad21 haploinsufficiency did result in a characteristic appearance of the FACS plot, with reduction of CD9 cell surface marker expression. Additionally, Rad21 mutant MkPs were more proliferative in culture than WT equivalents, and showed a trend towards reduced ploidy, which would be consistent with the known important role of cohesin in mitosis.

In general terms, combined Gata1s/Rad21 mutant mice displayed a more pronounced haematopoietic phenotype than either Gata1s or Rad21 mutant mice alone, supporting a role for mutational cooperativity. The most striking finding was expansion of the megakaryocyte progenitor compartment with an associated increase in bone marrow megakaryocytes and evidence of abnormal differentiation and proliferation behaviour of mutant MkP cells on *in vitro* culture. Interestingly, this was not accompanied by a change in platelet count in the peripheral blood - perhaps reflecting a degree of differentiation block with impairment in functional megakarypoiesis or some form of autoregulation.

When considering the potential processes underlying this phenotype, an important question concerns whether there is relative expansion of cells adopting a megakaryocyte fate, loss of repression of cells adopting an erthyroid (or other) fate or both. It has been shown that establishment of lineage fidelity involves both activation of lineage-specific gene expression patterns and repression of genes involved in alternative cell fates or earlier stages of development through the combinatorial action of transcription factor networks (Hong et al. 2005; Miccio et al. 2010; Rodriguez et al. 2005). Gata1 has been shown to function both as a transcriptional activator and repressor in megakaryo-erythropoiesis (Crispino 2005; Welch et al. 2004), acting in combination with other key proteins such as the Zinc finger protein FOG-1 (Crispino et al. 1999; Fox et al. 1999; Hong et al. 2005) and the NuRD co-repressor complex (Miccio et al. 2010). A consideration of

how Gata1s and a cohesin gene mutation may act to perturb these transcriptional processes is clearly of relevance.

Significantly, despite expansion of a relevant cellular population, the combined Gata1s/cohesin mutant mice did not develop leukaemia during the course of this study (including a cohort of mice which were followed for 1 year. Data not shown). This suggests that, at least in a mouse model system, Gata1s plus cohesin are insufficient for cellular transformation in the absence of Trisomy 21. Similar results were reported recently by Labuhn et al. (2019) who used a multiplex CRISPR-Cas9 loss-of-function screen in a Gata1s mutant fetal liver based murine model to assess the leukaemogenicity of recurrently mutated ML-DS genes. Loss of function of 18/22 recurrently mutated ML-DS genes resulted in a megakaryocytic leukaemia closely resembling ML-DS. However, cohesin gene mutations were the one group under-represented in this loss-of-function screen – perhaps suggesting the importance of a human trisomic background to disease pathogenesis in cohesin mutant cells.

So how do Gata1s and a haploinsufficient Rad21 mutation cooperate to cause MkP expansion? My hypothesis was that Gata1s and a reduced level of Rad21 protein synergise to alter gene expression programmes and therefore expand the MkP compartment - either directly or via suppression of alternative cell fate programmes. Gata1s may do this through direct target gene activation, while reduced Rad21 protein may alter chromosomal topology affecting expression of genes that act in the same pathway(s) as those regulated by Gata1s. I next addressed this hypothesis by performing molecular analysis of sorted MkPs.

5

Molecular analysis of the megakaryocyte progenitor population in *Gata1s* and cohesin mutant mice

Contents

5.1 Introduction

In the transgenic mouse model described in the previous chapter, the combination of a Gata1s mutation together with haploinsufficiency of the cohesin component Rad21 caused significant expansion of the bone marrow megakaryocyte progenitor compartment, with evidence of abnormal MkP proliferation and differentiation. To investigate how these mutations cooperate at a molecular level to drive this phenotype, I next looked at their impact on gene expression programmes in the MkP compartment.

5.1.1 Molecular regulation of megakaryopoiesis

Megakaryopoiesis is regulated by a variety of interrelated mechanisms, including the interaction of cytokines with cell surface receptors. The most critical of these cytokines is thrombopoietin (TPO), which together with its receptor c-mpl is important to all aspects of megakaryocyte development, including differentiation of megakaryocyte progenitor cells from the HSC (Bartley et al. 1994, Sauvage et al. 1994, Sohma et al. 1994, Kaushansky et al. 1994). TPO signalling results in internalisation of the bound c-mpl:TPO complex and initiation of multiple intracellular pathways, including JAK/STAT, PI3-Kinase and MAP-Kinase. These in turn cause activation of megakaryocyte-specific transcription factors and initiation of megakaryocyte specific gene expression programmes. Other cytokines, including IL-3, IL-6, IL-11 and SCF have also been shown to play significant roles (reviewed in Goldfarb 2009, Noetzli et al. 2019).

A number of transcription factors are important at different stages of megakaryocyte lineage commitment (Figure 5.1). These include factors which are specifically upregulated in the megakaryocytic lineage, such as RUNX and FLI1, and others, such as GATA1, GATA2, FOG1 and NF-E2 which are critical in a stage specific manner for both megakaryocyte and erythroid differentiation.

GATA1 and GATA2 are the major GATA factors expressed during megakaryocyte-erythroid development, with GATA2 levels decreasing and GATA1 levels rising during differentiation (Cantor et al. 2002). The GATA-co-regulator Friend of GATA 1 (FOG1) is co-expressed and stimulates or inhibits GATA1 activity depending on cellular context (X. Wang et al. 2002). It also recruits the NuRD repressor complex to facilitate GATA1-mediated transcriptional repression (Hong et al. 2005). Conditional deletion of *Zfpm1* (FOG1) and *Gata1* demonstrated that FOG-1 and Gata1 act sequentially, with FOG1 playing a role in the specification of PreMegE progenitors and Gata1 in the commitment of these progenitors to an erythroid fate (Mancini et al. 2012).

The transcription factor Erg (ETS-related gene) is essential for definitive haematopoiesis and the function of adult haematopoietic stem cells. It has

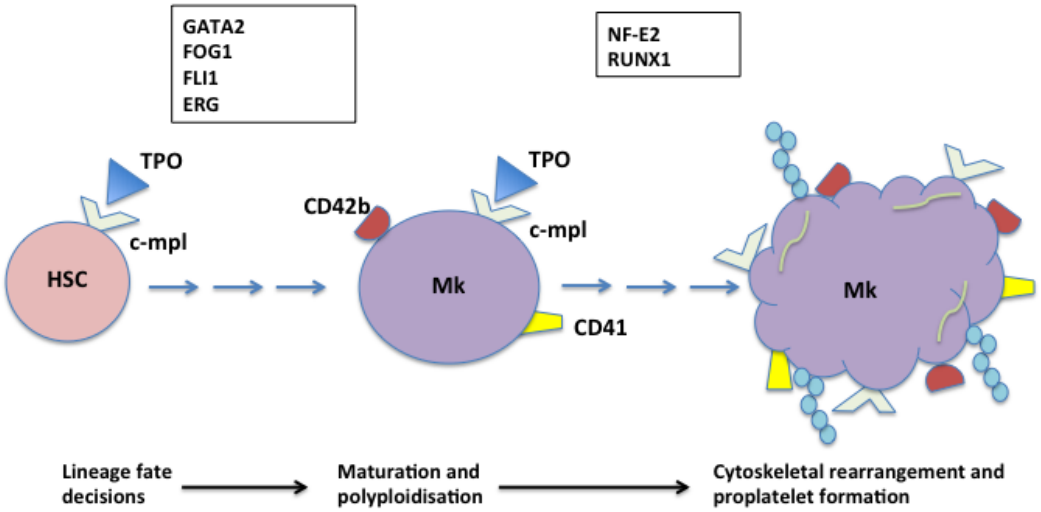


Figure 5.1: Key factors involved in the regulation of megakaryopoiesis. Schematic showing some of the cytokines and transcription factors involved in megakaryocyte differentiation. HSC = Haematopoietic stem cell; MK = megakaryocyte progenitor/mature megakaryocyte. CD41 and CD42b are cell surface markers upregulated during megakaryopoiesis. Blue circles represent proplatelets. Key transcription factors are boxed.

also been shown to play a key role in megakaryopoiesis and the maintenance of peripheral platelet count (Loughran et al. 2008). The closely associated ETS-family transcription factor FLI1 is important for expression of genes associated with terminal megakaryocytic differentiation (Starck et al. 2010). This has recently been confirmed in single cells at the protein level, by the finding that overexpression of FLI1 in bipotential E-MEG progenitors deviates cells towards a megakaryocytic fate (Palii et al. 2019). GATA-1 also antagonises another ETS-family transcription factor, PU.1 (a master myeloid regulator) through direct physical interaction to repress alternate lineage programmes (Nerlov et al. 2000, Chou et al. 2009). NF-E2 is critical for end stage megakaryopoiesis, controlling genes such as Beta1-tubulin and thromboxane synthase important for cytoskeletal rearrangement and proplatelet

formation (Shivdasani et al. 1995). RUNX1 represses the erythroid transcription factor KLF1, thus promoting megakaryocyte differentiation and suppressing erythroid differentiation (Kuvardina et al. 2015). During late megakaryopoiesis, it also directly regulates components of the megakaryocyte and platelet cytoskeleton (myosin heavy chain and myosin light chain) (Lordier et al. 2012). Ectopic expression of Klf1 suppressed both Fli-1 gene expression and megakaryocyte differentiation, while Klf1 knockdown promoted megakaryopoiesis (Frontelo et al. 2007).

It is increasingly recognised that cellular differentiation is not simply controlled by the action of individual transcription factors binding to cis-regulatory elements, but rather through simultaneous binding of transcription factor complexes in various combinations, with the interaction between factors determining an activating or repressive role on gene expression. Furthermore, local chromatin states and epigenetic modifications can also promote or limit transcription factor occupancies at specific loci, while 3D folding of the genome may bring distant cis-regulatory elements into contact, providing yet another level of transcriptional control (Denker et al. 2016).

5.1.2 Known roles of Gata1 and Rad21 in transcriptional regulation

Both Gata1 and Rad21 have been implicated in the regulation of eukaryocyte gene expression.

5.1.2.1 Gata1s and the regulation of gene expression

As discussed in Chapter 1.4, Gata1 is a key haematopoietic transcription factor required for the terminal differentiation of megakaryocytes (Shivdasani et al., 1997), erythrocytes (Fujiwara et al. 1996), mast cells (Migliaccio et al. 2003) and eosinophils (Yu et al. 2002).

Mouse models have provided considerable insight into the impact of full length, WT Gata1, and the N-terminally truncated form of Gata1, Gata1s, on gene expression in erythromegakaryopoiesis. Indeed, it has been shown that exclusive production of Gata1s perturbs gene expression profiles. In a transgenic knock in

mouse model exclusively expressing Gata1s, Li et al. observed a failure of Gata1s mutant megakaryocytes to repress expression of the transcription factors *Myc*, *Myb*, and *Gata2* (required for proliferation of haematopoietic progenitors) and *PU.1* and *Ikaros* (which play a role in specification of lineage choices other than megakaryocyte and erythroid) (Z. Li et al. 2005). In G1ME, a cell line derived from *Gata1*-deficient ES cells which has both erythroid and megakaryocytic differentiation potential, upon reconstitution of Gata1 expression, Chlon et al. revealed differential expression of 847 genes between Gata1 and Gata1s expressing cells. Many of the genes downregulated in Gata1s relative to Gata1 were genes specific to erythroid differentiation, such as *Hbax, Alas2, Eraf and Slc4a1*, suggesting that Gata1s may be less efficient at inducing the erythroid gene expression programme in this cellular context compared to full length Gata1 (Chlon et al. 2015). Similarly, in an induced pluripotent stem cell (iPSC) model derived from patients with Gata1s mutations, global transcriptome analysis revealed that Gata1s was associated with downregulation of erythroid genes and upregulation of myelo-megakaryocytic genes. Among those down regulated genes, 19 of 22 were likely Gata1 targets based on ChIP-Seq of human erythroblasts, suggesting that Gata1s may fail to efficiently bind or activate genes normally regulated by WT Gata1 (Byrska-Bishop et al. 2015). Finally, in a recently published study of purified embryonic erythroblasts from Gata1s mutant mice, Ling et al. reported significant dysregulation of Gata1 target genes, down regulation of genes associated with heme metabolism and haematopoietic cell maturation (*Klf1, Zfpm1, Lmo2 and Ldb1*) and enrichment for megakaryocytic pathways (Ling et al. 2019).

How Gata1s might differentially regulate gene expression remains incompletely understood. ChIP-seq, Cleavage Under Targets and Release using Nuclease (CUT&RUN) and Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq) in the above studies (Chlon et al. 2015, Byrska-Bishop et al. 2015, Ling et al. 2019) were used to compare Gata1FL to Gata1s binding. Taken together, these experiments suggest that Gata1 and Gata1s show relatively similar occupancy of Gata1 binding sites. However, some differences were noted. In G1ME

cells, genes assigned to sites bound less by Gata1s than Gata1fl were enriched for erythroid functions (Chlon et al. 2015, Byrska-Bishop et al., 2015). In Gata1s murine fetal erythroblasts, Gata1s binding and chromatin accessibility were gained at 329 sites and lost at 1844 sites compared to WT, with differential sites largely found in gene bodies and intergenic regions. This included enhanced binding at some megakaryocytic genes (*Fli1 and Mef2c*) and reduced binding at erythroid genes (*Lmo2 and Zfpm1*). *Gata2* and *Runx1* were expressed at moderately increased levels in Gata1s mutant cells with associated increased accessibility, suggesting failure of appropriate repression by Gata1s. This was accompanied by a loss of the repressive histone mark H3K27me3 at the *Gata2* locus, leading the authors to suggest a role for reduced recruitment of Polycomb Repressor Complex 2 (PRC2) to key loci in Gata1s mutant cells, although no direct interaction between Gata1 and Ezh2 or Suz12 was found (Ling et al. 2019).

5.1.2.2 Haploinsufficiency of cohesin and modulation of gene expression

In addition to its role in maintaining chromosome cohesion during mitosis, the cohesin complex is known to regulate both DNA replication and long-range regulation of gene expression via the formation and stabilisation of chromatin loops (Merkenschlager et al. 2016). Cohesin localisation is mediated in part by the transcriptional regulator CTCF (Wendt et al. 2008), although it can also associate with chromatin independently (Kagey et al. 2010). One theory is that via DNA looping, cohesin acts to regulate enhancer-promoter interactions within topology-associated domains (TADs) and chromosomal compartments, thus influencing gene expression (Denker et al. 2016), however it is increasingly apparent that this may not be a straightforward relationship.

Indeed, the changes in gene expression observed in cohesin deficient cells are often relatively subtle. Gene expression profiling of lymphoblastoid cell lines derived from patients with the cohesinopathy Cornelia de Lange syndrome, revealed around 400 dysregulated genes in mutant cells, typically with a less than two-fold change in level of expression compared to WT (Liu et al. 2009). Cohesin was

mainly bound to active transcriptional start sites and the authors suggested that the observed clinical phenotype might result from an accumulation of multiple small effects on gene regulation.

Similarly, recent studies investigating the effect of cohesin haploinsufficiency in haematopoietic cells in mouse and human models revealed increased expression of stem related genes and reduced expression of mature differentiation genes (Viny et al. 2015, Mazumdar et al. 2015, Mullenders et al. 2015, Viny et al. 2019). These gene expression changes were accompanied by a global decrease in chromosome accessibility (as assayed by ATAC-seq of mutant and WT bone marrow cells) but increased accessibility associated with motifs for the transcription factors GATA2, RUNX1 and ERG (Mazumdar et al. 2015). A role has also been suggested for cohesin in the dynamic remodelling of transcription during erythropoiesis, with depletion of cohesin *in vitro* associated with impaired erythroid differentiation and enhanced self renewal - particularly at loci bound by ETV6, where cohesin is hypothesised to ameliorate ETV6-mediated gene repression (Sasca et al. 2019).

Interestingly, the effect of cohesin depletion on TAD and sub-TAD structure appears to be variable. In post mitotic murine CD4+ lymphocytes, depletion of RAD21 resulted in weakening of intraTAD interactions but a strengthening of long range interTAD interactions and resulted in defective Tcra transcription (Seitan et al. 2011). By contrast, fusion of TADs at the Sox9-Kcnj2 locus was achieved only following removal of all CTCF sites (both within the TADS and at the boundary) (Despang et al. 2019) and did not result in any major changes in gene expression, suggesting that, at some loci at least, TADs may be dispensable for gene regulation.

5.1.3 Aims

Both Gata1s and cohesin haploinsufficiency alone have been shown to perturb gene expression. One hypothesis is that Gata1s does this via differential activation or repression of target genes, while depletion of Rad21 alters chromosome topology and causes subtle dysregulation of multiple loci. Asking how these two mutations

synergise in haematopoietic cells at a molecular level to expand the MkP compartment is relevant - both to understanding the described phenotype, and also considering how co-operating mutations contribute to the development of leukaemia. Therefore the aims for this section of work were:

1. To characterise differences in the transcriptional profiles of the MkP compartment between: (i) WT; (ii) *Gata1s*; (iii) *Rad21* haploinsufficient and (Iv) Combined *Gata1s/Rad21* mutant mice.

2. To optimise low cell number ATAC-seq protocols and use these to assess the extent of differences in chromatin accessibility.

3. To identify perturbed pathways which may lead to the observed expansion of MkP.

5.2 Results

5.2.1 RNA sequencing of sorted megakaryocyte progenitors

5.2.1.1 Processing and quality control

RNA-sequencing was performed on MkP (Lin- Sca1- CD16/32low CD41+ Kit+ CD9+) cell populations sorted from WT (n-4), Rad (n=3), Gata (n=4) and Comb (n=3) mutants. Details of sorting strategy, library preparation, sequencing and data analysis are outlined in Chapter 2 (Materials and Methods). Filtered trimmed reads were mapped to the mm10 genome with an efficiency of >80% uniquely mapping reads (Figure 5.2). Read counts for each annotated gene were quantitated using featureCounts (Liao et al. 2014) and gene expression level was calculated as Transcripts Per Million (TPM).

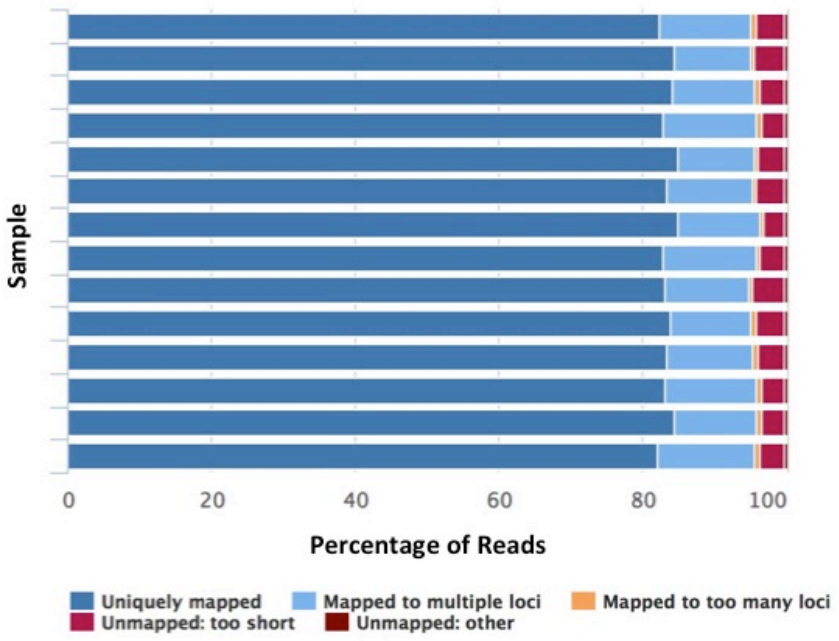


Figure 5.2: Mapping statistics of sorted MkP samples A summary of mapping statistics of bulk RNA-seq MkP populations. Mapping performed using STAR. Plots generated using Multiqc (Ewels et al. 2016). Each row represents one RNA-seq sample.

5.2.1.2 Overview of gene expression data: TPM scores

The total number of expressed genes was defined as TPM >1 and ranged from 22,383-22,484 per sample. The most highly expressed genes found in WT samples (Top 50 genes shown in Appendix A.3) were consistent with expected MkP transcriptional activity – including genes such as *PF4* (platelet factor 4), *B2M* (beta-2-microglobluin), *Calm2* (calmodulin 2) and *CD9*. EnrichR enrichment analysis (Chen et al. 2013) of the top 5000 most highly expressed genes in WT samples showed significant enrichment for the mega-erythroid progenitor gene set from the curated mouse gene atlas dataset (Su et al. 2004) (Figure 5.3)

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	mega_erythrocyte_progenitor	1.219e-40	1.170e-38	2.03	186.67
2	embryonic_stem_line_Bruce4_p13	1.322e-26	6.344e-25	1.64	97.74
3	embryonic_stem_line_V26_2_p16	6.894e-20	2.206e-18	1.60	70.63
4	adipose_brown	6.631e-14	1.591e-12	1.63	49.40
5	bone_marrow	6.733e-13	1.293e-11	1.63	45.77
6	Baf3	1.182e-8	1.890e-7	1.48	26.99
7	stem_cells__HSC	0.000003113	0.00004269	1.61	20.41
8	heart	0.00001204	0.0001445	1.31	14.88
9	follicular_B-cells	0.00005725	0.0006107	1.28	12.48
10	granulo_mono_progenitor	0.0005278	0.004606	1.54	11.66

Figure 5.3: Enrichment analysis of top 5000 most highly expressed genes in WT Megakaryocyte Progenitors. Table generated using the EnrichR package (Chen et al. 2013). Enrichment for upregulated genes in mouse tissues as reported in the Mouse Gene Atlas (Su et al. 2004).

Consistent with the observed decrease in Rad21 protein level in Rad21fl/+ VavCre tg mutant mice (Figure 4.3c), *Rad21* gene expression was reduced in Rad and Comb mutants compared to WT and Gata mice (Figure 5.4). In keeping with the phenotype described in Chapter 4, there was also a decrease in *CD9* gene expression in Rad and Comb mutant samples. Combined mutants displayed reduced expression of genes associated with megakaryocyte maturity *(PF4, VWF)* with a trend towards increased expression of more immature markers *(CD34, MPL)*.

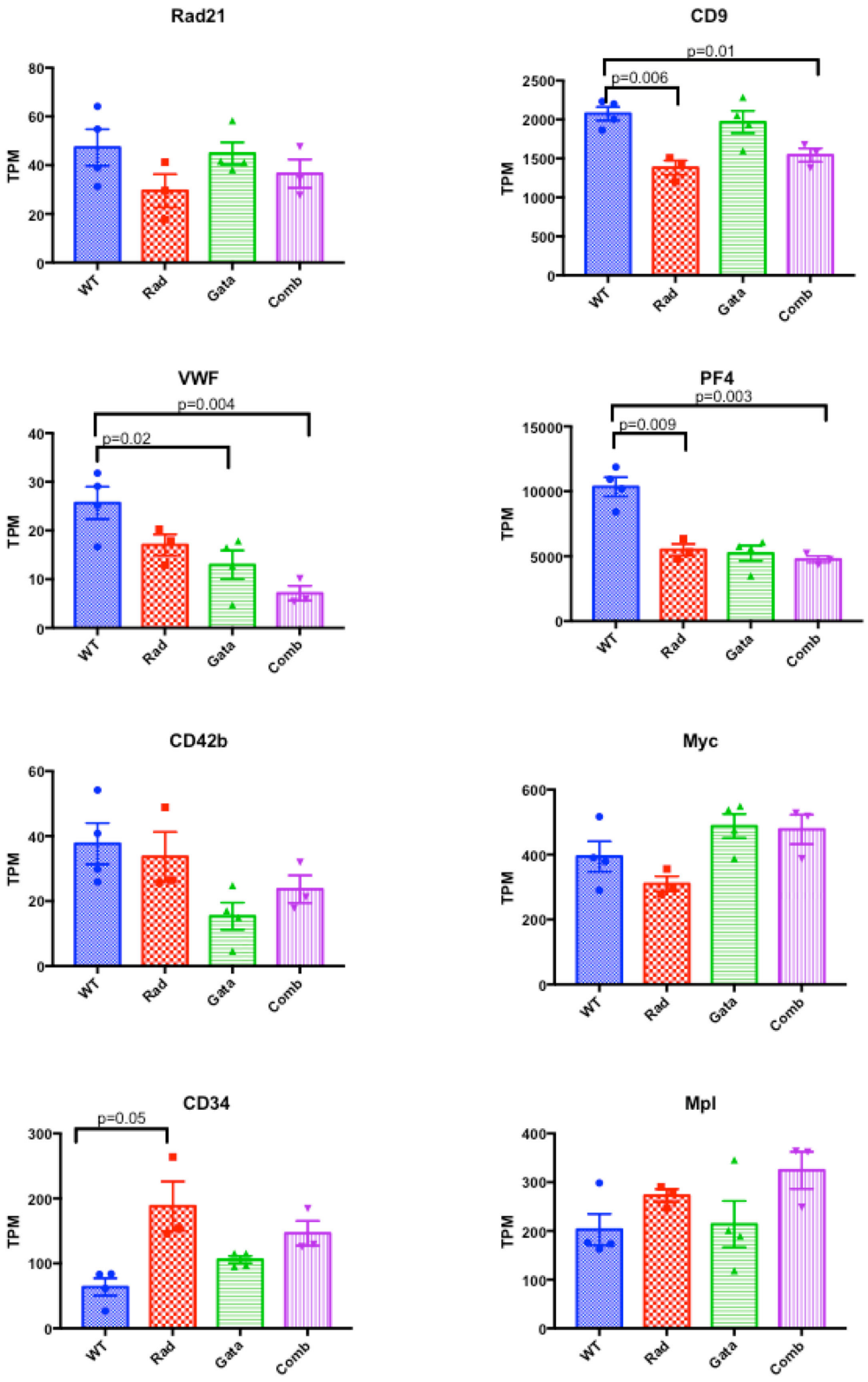


Figure 5.4: Gene expression in Megakaryocyte Progenitors. Graphs show Transcript Per Million (TPM) for WT, Rad, Gata and Cohesin mutant mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

I also looked specifically at the expression of key transcriptional regulators of megakaryopoiesis (Figure 5.5). Expression of *Gata2* was significantly increased in the Combined mutant compared to WT (p=0.02, One-way ANOVA). A moderate increase in *Gata2* expression has been previously reported in Gata1s mutant fetal liver, G1ME and erythroid cells (Li et al. 2005, Chlon et al. 2015, Ling et al. 2019) and postulated to result from reduced Gata1-mediated repression of *Gata2* expression (although the exact mechanism remains unclear). Interestingly, my data suggest a trend towards increased MkP *Gata2* expression in the combined mutant greater than that resulting from Gata1s alone, although this comparison does not reach significance by One-way ANOVA. A similar pattern of expression across genotypes was seen for the transcription factor *NF-E2* (p=0.009, One-way ANOVA), which is important for terminal maturation of megakaryocytes. This is in keeping with the observation that although Comb mutant MkPs show abnormal *in vitro* proliferation and differentiation, *in vivo* they are able to differentiate into mature platelet-producing megakaryocytes as shown by bone marrow trephine and platelet count. As expected, expression of the erythroid marker *Klf1* was significantly decreased in Gata samples and this remained low in the Comb mutant. Finally, the expression of the megakaryocytic transcription factor *Runx1* was significantly decreased in the combined mutant compared to WT (p=0.03, One-way ANOVA) - a finding is in keeping with a gene expression data from ML-DS patient samples, in which *RUNX1* was downregulated when compared to non-Down syndrome acute megakaryoblastic leukaemia (Bourquin et al. 2005).

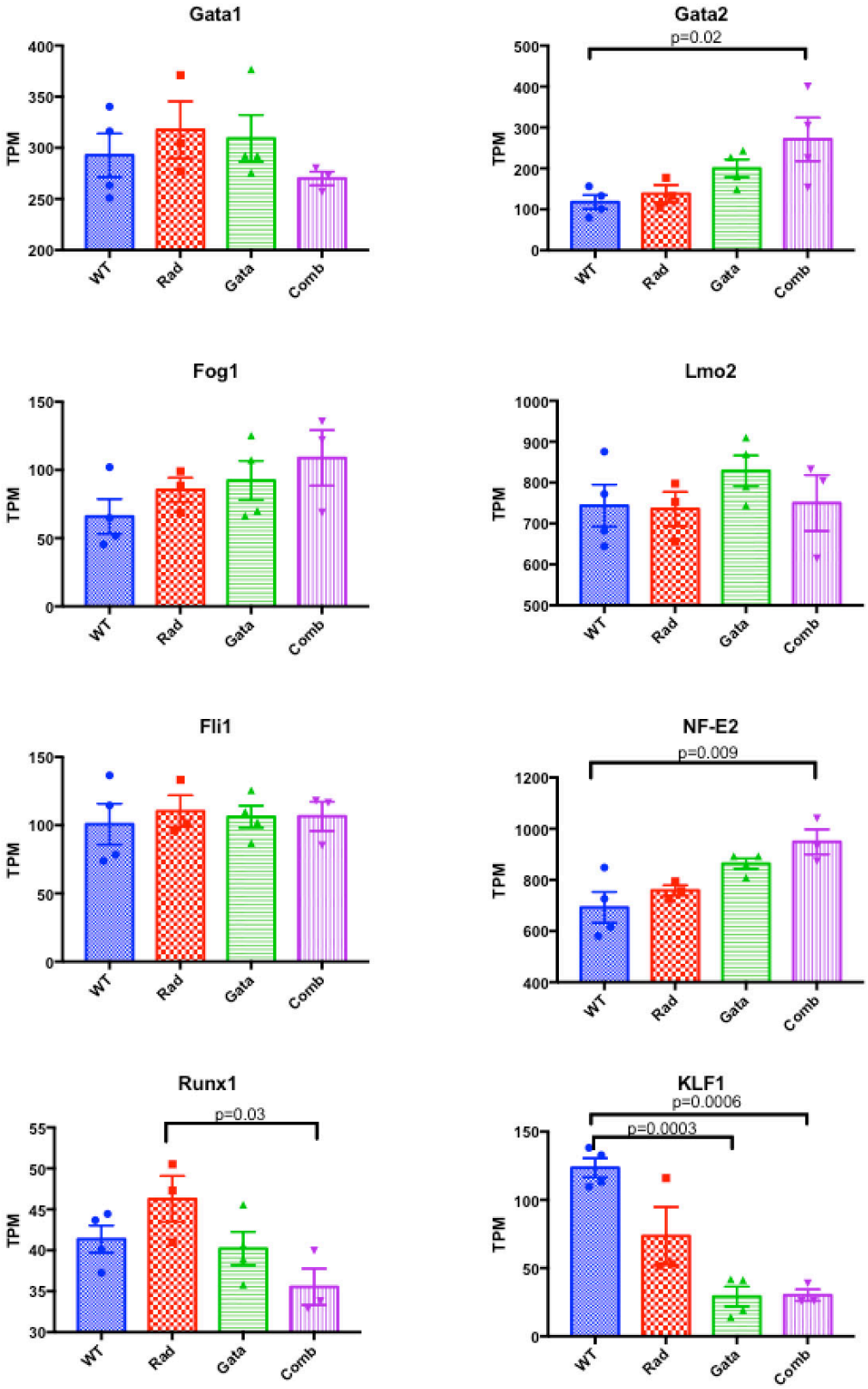


Figure 5.5: Expression of transcription factors important in megakaryopoiesis. Graphs show Transcript Per Million (TPM) for WT, Rad, Gata and Cohesin mutant mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown)

5.2.1.3 Principal component analysis

The global gene expression profile of sorted WT, Rad, Gata and Comb MkP cells was assessed in an unbiased manner using principal component analysis (PCA) of all expressed genes (Figure 5.6).

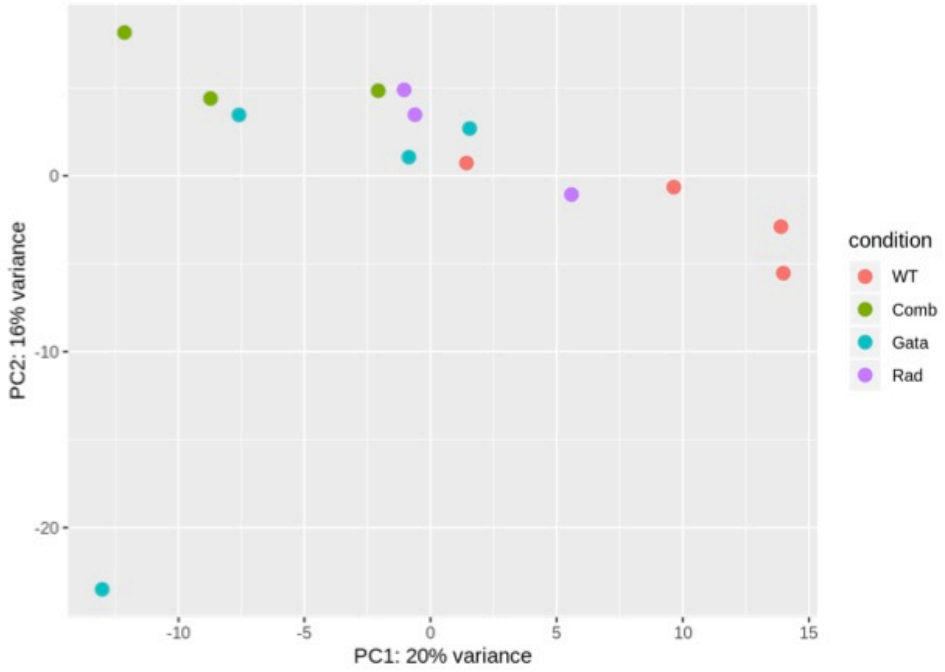


Figure 5.6: PCA of Megakaryocyte Progenitors using all expressed genes. The percentage variance contributed by each principal component is indicated. Each dot represents a biological replicate.

PC1, representing the majority of the variance, separated WT (orange) and Comb (green) samples, although gave relatively poor distinction between Rad (purple) and Gata (blue). PC2 was mainly driven by one biological replicate of Gata. To gain better distinction between groups, a second PCA was carried out using an unbiased selection of the 500 genes with the highest variance across all samples (Figure 5.7).

It was now possible to see segregation of the four genotypes, with PC1 (24% of the variance) broadly separating Gata1s mutant (Gata and Comb) from Gata1 wild type

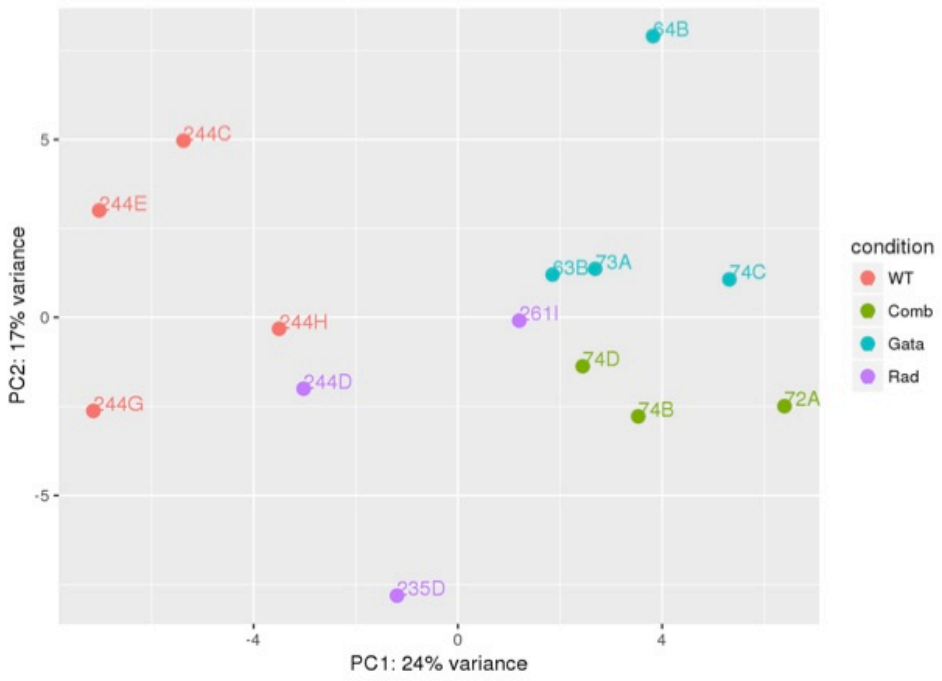


Figure 5.7: PCA of Megakaryocyte Progenitors using top 500 most variable genes. The percentage variance contributed by each principal component is indicated. Each dot represents a biological replicate.

samples (WT and Rad) and PC2 (17% of the variance) broadly separating Rad21 mutant samples (Rad and Comb) from Rad21 wild type samples (WT and Gata).

In order to investigate the principal components representing meaningful biological variance, the eigenvalues for the second PCA (Figure 5.7) were calculated. As shown in Figure 5.8, most of the biological variation was captured in PC1-4. Visualisation of PC3 and PC4 did not help to segregate samples further. All samples were processed in the same sequencing run. No batch effect was seen on PCA when samples were labelled by cDNA tagmentation batch or VavCre status (data not shown).

To investigate the genes underlying the principal components identified, I extracted the loading scores for PC1 and PC2 and annotated the top 50 contributing genes on a PCA plot (Figure 5.9).

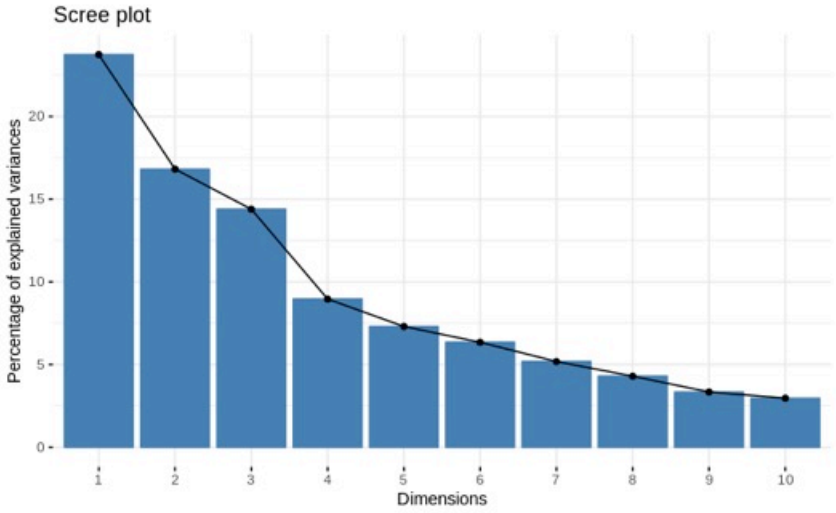


Figure 5.8: Scree plot showing contribution of each principal component to the variance. Dimensions represent PC1-PC10.

In the direction of the WT samples (shown here in blue) were a number of genes relevant to megakaryopoietic differentiation, including *PF4*, *VWF* and *Timp3*. Conversely, genes associated with less mature haematopoietic cells (*Gata2, CD34, Itga2b, MPO*) appeared to be driving the Gata (green) and Comb (yellow) clusters.

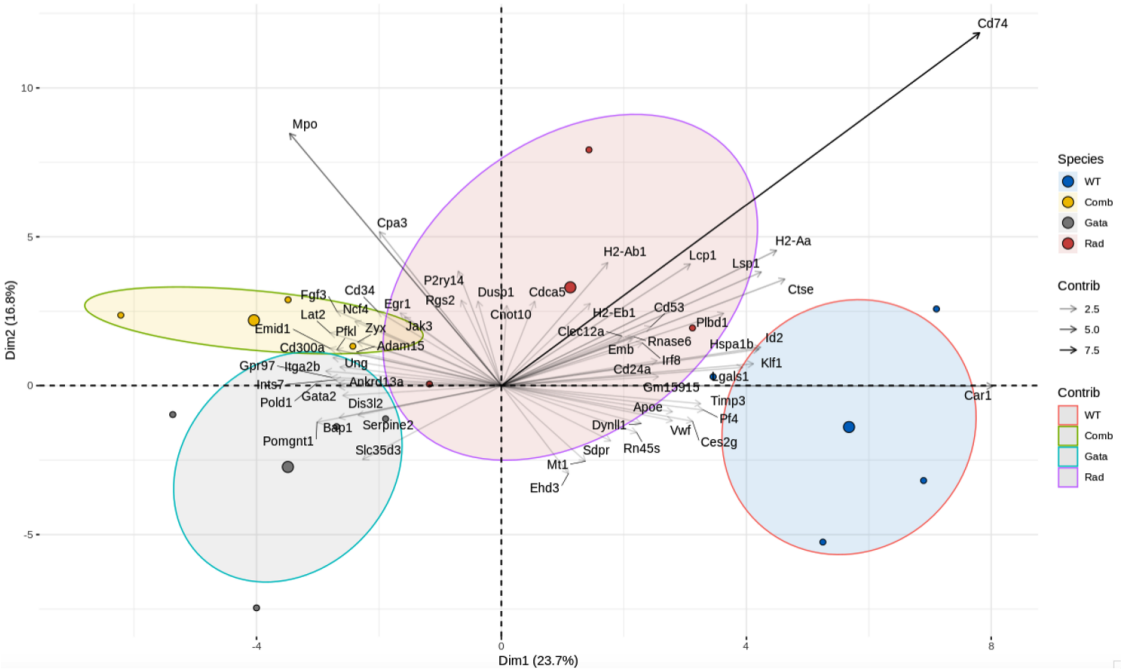


Figure 5.9: Loadings plot showing genes contributing to PC1 and PC2. Dim1 = PC1, Dim2 = PC2. Samples are labelled by colour. Shaded ellipses represent 95% confidence intervals around identified sample clusters. Each small dot represents a biological replicate. Large dots represent the cluster mean.

5.2.1.4 Differential gene expression

Differentially expressed genes were identified between control and mutant samples using Wald test pairwise comparisons in DESeq2 (Love, Huber and Anders, 2014). The number of significantly (p-adjusted value <0.05) differentially expressed genes in each comparison is shown in Figure 5.10a and represented in the MA (minus over average) plots in Figure 5.11.

a

	Total	Up	Down
Comb vs WT	665	252	413
Gata vs WT	151	92	59
Rad vs WT	56	31	25
Comb vs Gata	10	4	6
Comb vs Rad	23	9	14
Gata vs Rad	44	26	18

b

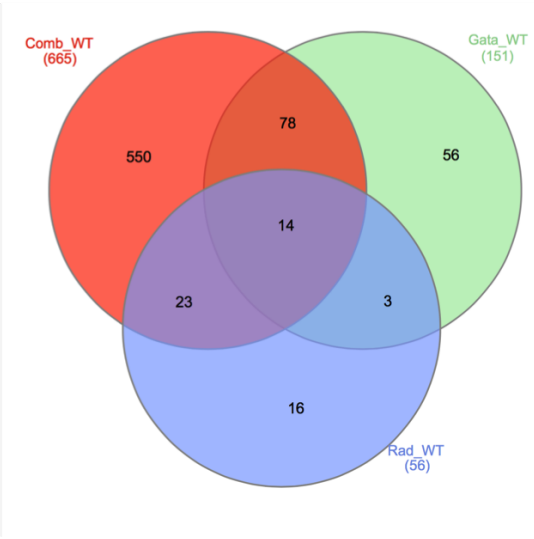


Figure 5.10: Differentially expressed genes (p-adjusted value <0.05). (a) Summary of significantly differentially expressed genes between genotype groups. Total = total number of differentially expressed genes with p-adjusted value <0.05; Up = genes upregulated in first of pair compared to second; Down = genes down regulated in first of pair compared to second. (b) Venn diagram showing overlap between differentially expressed gene lists for Comb vs WT, Rad vs WT and Gata vs WT pairs. Gene names, p-adjusted values and log2fold change for each gene set can be found in Appendix A.4. Differentially expressed gene lists were identified by Wald Test pairwise analysis in DESEQ2.

In the MA plots (Figure 5.12), log Fold change (logFC) is plotted against normalised gene expression, with a positive fold change indicating upregulation in the first genotype of the pair. In keeping with the magnitude of the described

phenotypic effect (Figure 4.19), the greatest number of significantly differentially expressed genes was found when comparing Comb to WT and the smallest number between Comb and Gata. For all comparisons, absolute numbers of differentially expressed genes were relatively small - perhaps reflecting the fact that analysis was performed on a highly sorted population of cells. Full lists of gene names, p-adjusted values and log2fold change can be found in Appendix A.4. The number of overlapping genes differentially expressed in the Comb vs WT, Gata vs WT and Rad vs WT comparisons are shown in Figure 5.11b. The greatest degree of overlap was seen between the Comb vs WT and Gata vs WT gene lists, reflecting the relative importance of mutant Gata1s to the observed phenotype.

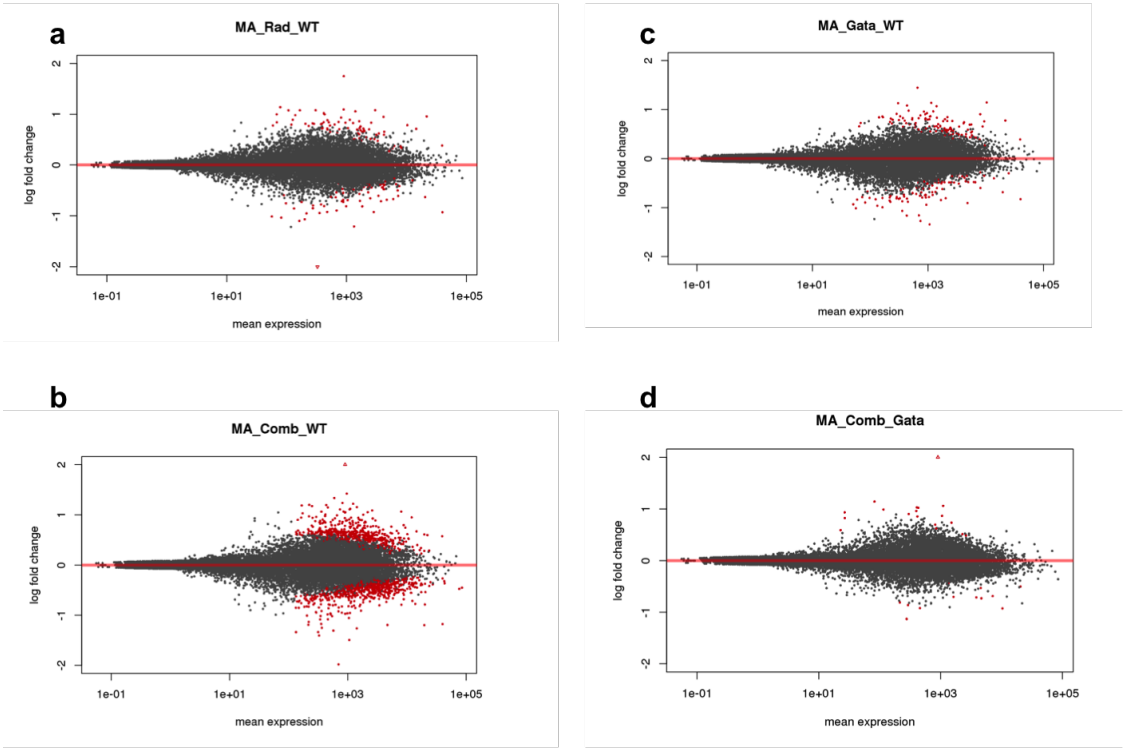


Figure 5.11: MA plot showing differentially expressed genes between genotypes. Differentially expressed genes between Rad and WT (a), Comb and WT (b), Gata and WT (c) and Comb and Gata (d) were identified using a Wald test. Mean expression represents the mean expression of each gene. Positive log fold change reflects upregulation in the first of the pair compared to the second and negative log fold change reflects down regulation in the first of the pair compared to the second. Red dots represent genes with a p-value <0.1.

5.2.1.5 Normalisation of sequencing data using ERCC spike in controls

It has previously been suggested that haploinsufficiency of cohesin may lead to global changes in gene expression. Viny et al. reported a global decrease in RNA transcript levels in Kit+ bone marrow cells from Smc3+/- mice compared to wild type controls (Viny et al. 2015). This was only detected by the inclusion of an exogenous spike-in control to allow for cell number normalisation. Global changes in gene expression were not observed in other cohesin deficient mouse studies (Mazumdar et al. 2015, Mullenders et al. 2015), although these studies did not utilise spike-in controls.

To try and detect any global changes in gene expression caused by cohesin deficiency, I added External RNA Controls Consortium (ERCC) Spike-in controls (Life Technologies) to sorted cells prior to cDNA synthesis. ERCC Spike-ins are a set of unlabelled, polyadenylated transcripts designed to mimic eukaryoctic mRNAs of known concentrations. These can then be used for normalisation between samples. In my dataset, the relationship between mapped ERCC reads and nominal concentration was broadly linear (Figure 5.12, R=0.78). However, as reported by others (Risso et al. 2014) the proportion of reads mapping to ERCC spike ins varied considerably between samples and in some cases differed markedly from the expected value. This variation did not appear to be influenced by genotype and might reflect technical variation between samples. For the purposes of normalisation, two outlying samples were removed prior to downstream analysis. The remaining samples were processed using the RUVg pipeline from RUVseq (Removal of Unwanted Variation in RNA sequencing data) (Risso et al. 2014), which uses negative control genes (in this case ERCC spike-ins) assumed to have a constant expression across samples, to normalise data. The RUVg output was then analysed as above using DESEQ2 to generate differential gene lists.

As shown in the MA plots in Figure 5.13, no significant change in differentially expressed gene lists or the proportion of up and down regulated genes in cohesin mutant samples was detected following normalisation to ERCC spike-in controls. Although these data argue against a global change in gene expression, accurate normalisation remains an important issue and is discussed in more detail below.

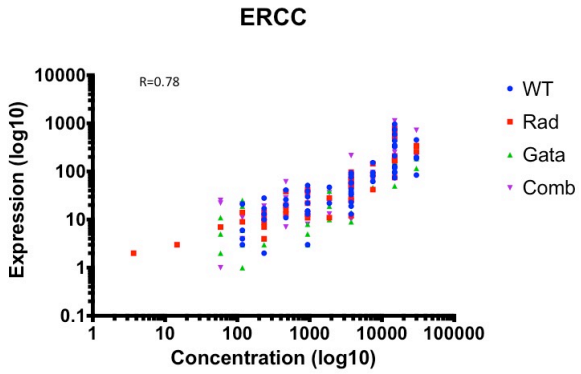


Figure 5.12: Relationship between mapped ERCC spike-in controls and nominal concentration for all samples is linear but shows inter-sample variability. Expression = number of mapped reads (log scale). Nominal concentration calculated using manufacturer's guidebook (Life Technologies).

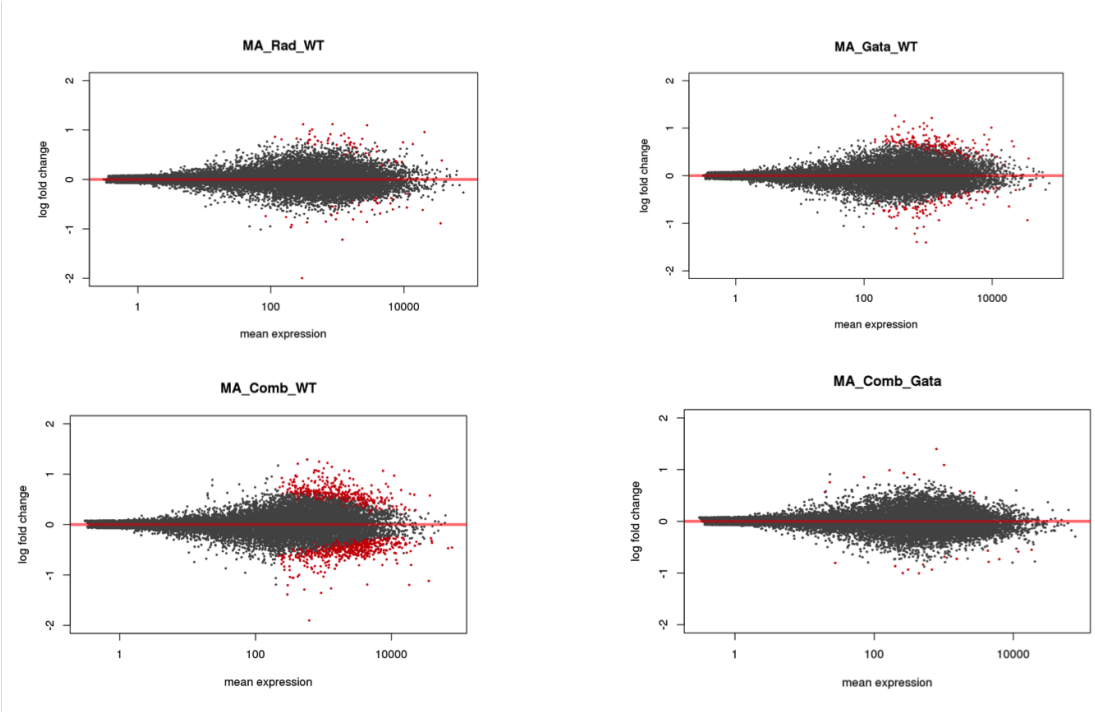


Figure 5.13: MA plot showing differentially expressed genes between genotypes following normalisation using ERCC spike-in controls. Samples were normalised using RUVg-seq and differentially expressed genes between Rad and WT (a), Comb and WT (b), Gata and WT (c) and Comb and Gata (d) identified using a Wald test. Mean expression represents the mean expression of each gene. Positive log fold change reflects upregulation in the first of the pair compared to the second and negative log fold change reflects down regulation in the first of the pair compared to the second. Red dots represent genes with a p-value <0.1.

Given the possible limitations of these experiments in view of the variation between samples and need to exclude replicates on technical grounds, ongoing analysis focuses on the gene lists generated without ERCC-based normalisation.

5.2.1.6 Gata1s and cohesin gene mutations together alter gene expression profiles: Analysis of the gene signature of combined mutant mice

As shown in Figure 5.10, 151 genes were significantly differentially expressed between the Gata1s mutant and WT. Consistent with previous studies (Chlon et al. 2015, Byrska-Bishop et al. 2015, Ling et al. 2019), and the observed *in vitro* phenotype, Gata1s mutant MkP cells demonstrated negative enrichment for genes associated with terminal erythropoiesis compared to the WT and positive enrichment for known Gata1 target genes (Encode ChIP and CHIA-PET data, p-adjusted value 0.007).

A smaller number of genes (56) were significantly differentially expressed between Rad21 haploinsufficient and WT samples (p-adjusted value <0.05). Interestingly, these included *CD9, PF4* and *CD34* and enriched for a number of signalling pathways, including TNF-alpha and IL-2 via STAT5.

By contrast, the combination of Gata1s and haploinsufficiency of Rad21 resulted in differential expression of a much larger number of genes (665) in sorted MkP cells compared to WT controls - much greater differential gene expression than either Gata1s or Rad21 mutations alone. This strongly suggests a synergistic (or additive) effect on gene expression programmes leading to expansion of the MkP.

The significantly differentially expressed genes between combined mutant and WT samples were visualised in a heatmap (Figure 5.14), annotated here with all differentially expressed genes with a p-adjusted value <0.05 and Log2FC >1 or <-1. This revealed a number of relevant loci, including *Timp3, Klf1, Vwf, Pf4, Sdpr, Mta3, CD34* (Log2FC 0.78) and *Gata2* (Log2FC 0.68). Interestingly, it was noted that a number of the genes differentially expressed between Comb and WT mice were located in adjacent chromosomal positions, perhaps suggesting a role for perturbed boundary elements and/or chromosomal structure in their dysregulation. Examples include *vWf* and *CD9*, *P2ry14* and *Gpr171* and *MTA3* and *Haao.*

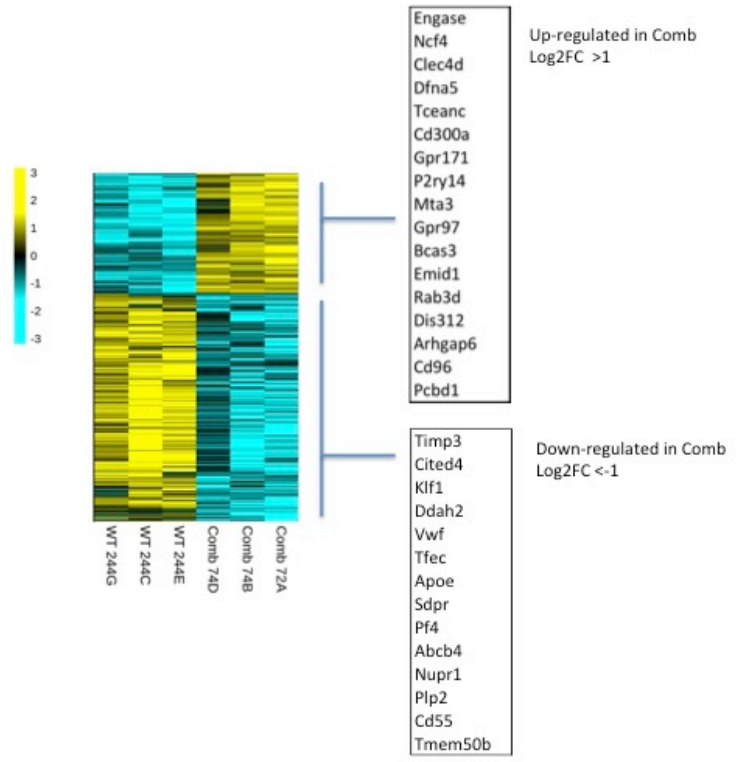


Figure 5.14: Gata1s plus a cohesin mutation perturb gene expression:heatmap of significantly differentially expressed genes in the MkP of Combined mutant vs WT. Differential gene list identified using a Walt Test in DESEQ2 with p-adjusted value <0.05. Columns represent biological replicates of WT (left) and Comb (right) samples. Rows represent genes. Colour scale is normalised per row with scale shown on left. For full list of gene names, p-adjusted values and log2fold change see Appendix A.4

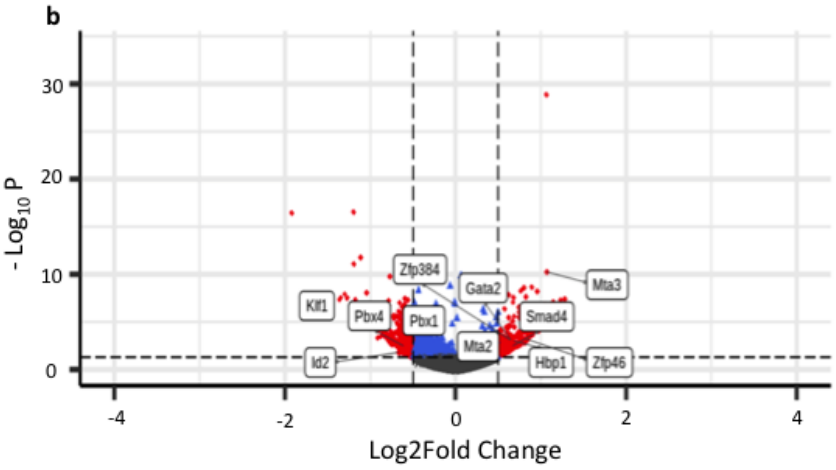
The differentially expressed gene list was then intersected with a list of known murine transcription factors from Encode (Davis et al. 2018). The results are summarised in Figure 5.15 and include a number of transcription factors with known Gata1 binding sites, including Klf1, MTA2 and MTA3 (NuRD complex), Gata2, Id2 and Pbx4. Consistent with this, the list of differentially expressed transcription factor genes was significantly enriched for Gata1 binding (p=0.004, EnrichR analysis (Chen et al. 2013)).

Enrichment analysis was performed on the list of significantly differentially expressed genes between Combined mutant and WT MkPs using both EnrichR (Chen et al., 2013) and Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005). This revealed enrichment in the combined mutant for upregulation of KEGG

a

TF	Log2FC
Klf1	-1.352973125
Tfec	-1.201130027
Pbx1	-0.683884534
Id2	-0.673027443
Pbx4	-0.619652367
Nme2	-0.448709641
Lyar	-0.402763798
Elf1	0.417743546
Mkrn2	0.488169961
Rit1	0.516661404
Mta2	0.556621158
Zfp384	0.638989455
Gata2	0.647618154
Smad4	0.675362584
Hbp1	0.688117656
Zfp46	0.833323975
Mta3	1.069742756

b



c

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	GATA1_CHEA	0.004112	0.4276	5.83	32.04
2	NFE2L2_CHEA	0.009455	0.4916	4.60	21.46
3	NFIC_ENCODE	0.2125	0.9209	4.22	6.53
4	BCL3_ENCODE	0.2084	0.9425	4.31	6.76
5	AR_CHEA	0.2379	0.9517	2.15	3.09
6	SP2_ENCODE	0.2059	0.9731	2.37	3.74
7	SALL4_CHEA	0.2626	0.9753	3.31	4.43
8	MAX_ENCODE	0.2553	0.9835	1.70	2.32
9	EGR1_CHEA	0.2366	0.9843	3.73	5.38
10	FOXA2_ENCODE	0.02894	1.000	7.45	26.38

Figure 5.15: Gata1s plus a cohesin gene mutation perturbs expression of key transcription factors. (a) List of transcription factors (TF) upregulated (pink) and downregulated (blue) in the MkP of Combined mutant mice compared to WT together with associated Log2Fold change (Log2FC). (b) Volcano plot showing the -log10 of the p value against the log2fold change for genes differentially expressed between the MkP of Combined mutant and WT mice. Each dot represents a gene. Red = log2Fold change >0.6 or <-0.6 and p-adjusted value <0.05. Blue = p-adjusted value <0.05. Gene list identified using Wald test in DESEQ2 and full list of gene names, log2fold change and p-adjusted values in appendix. (c) Enrichment analysis of differentially expressed transcription factor gene list (a) shows enrichment for Gata1 binding. Table of top 10 most enriched transcription factor pathways generated using the EnrichR package (Chen et al. 2013) based on consensus binding from Encode and ChEA data sets.

pathways associated with cell proliferation (DNA replication (P=0.0004), Cell cycle (P= 0.002) and glycolysis/gluconeogenesis (P-0.005)) (Figure 5.16, upper panel). The upregulated gene list was also enriched for binding by key haematopoietic transcription factors including RUNX1, KLF4, SPI1, GATA2 and E2F4 (Figure 5.16, lower panel).

KEGG 2019 Mouse

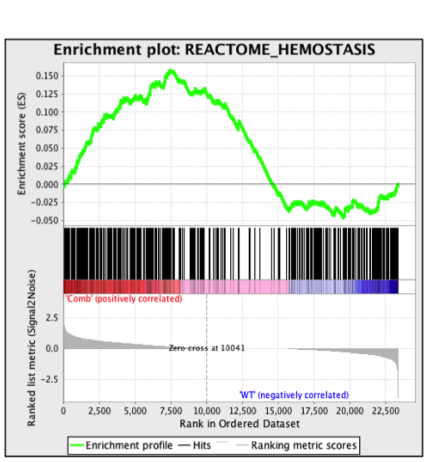
Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	DNA replication	0.0004966	0.1505	10.78	82.02
2	Cell cycle	0.002008	0.3043	4.60	28.58
3	Hematopoietic cell lineage	0.003266	0.3298	5.02	28.72
4	Central carbon metabolism in cancer	0.004746	0.3595	5.90	31.55
5	Glycolysis / Gluconeogenesis	0.005587	0.3385	5.63	29.22
6	Leishmaniasis	0.005587	0.2821	5.63	29.22
7	Primary immunodeficiency	0.006481	0.2805	7.86	39.61
8	Platelet activation	0.01074	0.4066	3.77	17.11
9	Osteoclast differentiation	0.01181	0.3976	3.69	16.36
10	Nitrogen metabolism	0.01370	0.4150	11.10	47.62

ENCODE and ChEA Consensus TFs from ChIP-X

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	RUNX1_CHEA	0.000002370	0.0002465	2.41	31.16
2	KLF4_CHEA	0.00001796	0.0009340	2.49	27.16
3	SPI1_ENCODE	0.0008779	0.03043	1.84	12.93
4	SPI1_CHEA	0.001954	0.05081	1.97	12.26
5	GATA2_CHEA	0.003651	0.07594	2.08	11.66
6	E2F4_ENCODE	0.003816	0.06615	2.13	11.84
7	CHD1_ENCODE	0.004354	0.06468	2.16	11.75
8	FOXA1_ENCODE	0.006354	0.08260	3.22	16.30
9	E2F6_ENCODE	0.008977	0.1037	1.40	6.58
10	IRF3_ENCODE	0.01137	0.1182	1.99	8.92

Figure 5.16: Datasets enriched in Combined mutant versus WT control. Tables of top 10 most enriched transcription factor pathways generated using the EnrichR package (Chen et al. 2013) based on KEGG (2019 Mouse) pathways analysis and consensus binding from Encode and ChEA data sets.

GSEA analysis revealed positive enrichment for genes associated with early megakaryopoiesis (ITGA2b (CD41), MPL, GATA2) in Comb compared to WT MkP but negative enrichment for genes associated with more mature megakaryocytes (VWF, PF4, CD9) giving a biphasic curve (Figure 5.17). Gene sets annotated to cell cycle (REACTOME), DNA replication (REACTOME), E2F Targets (HALLMARK) and Gata2 targets (Huang et al., 2009) were also significantly enriched (Figure 5.18, FDR <0.25).



GeneSet	REACTOME_HEMOSTASIS
Enrichment Score (ES)	0.15832207
Normalized Enrichment Score (NES)	3.785993
Nominal p-value	0.0
FDR q-value	0.0

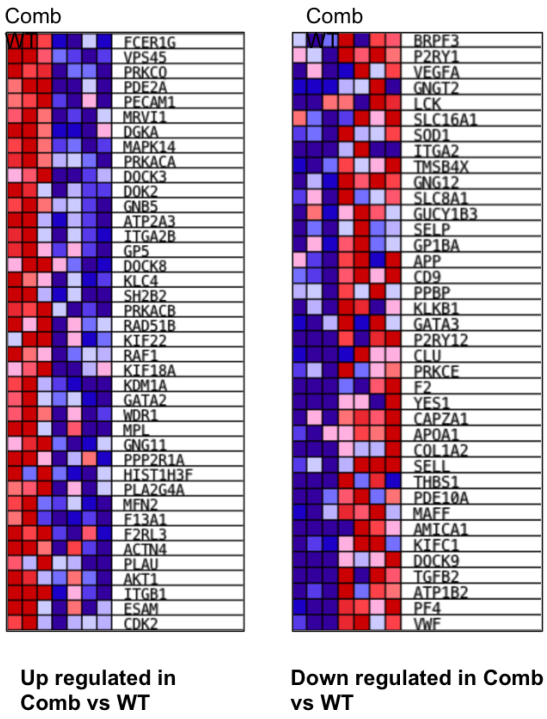
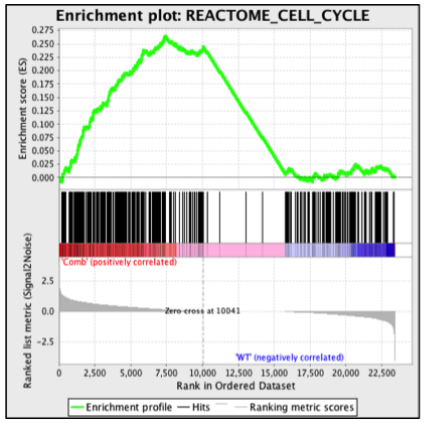
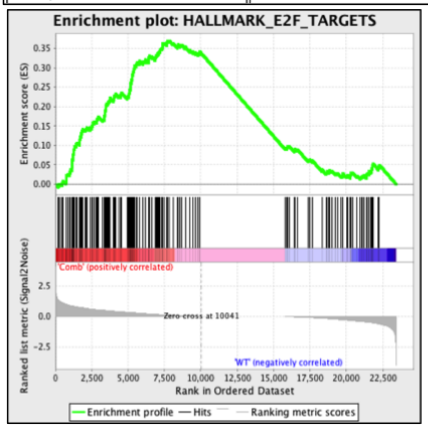


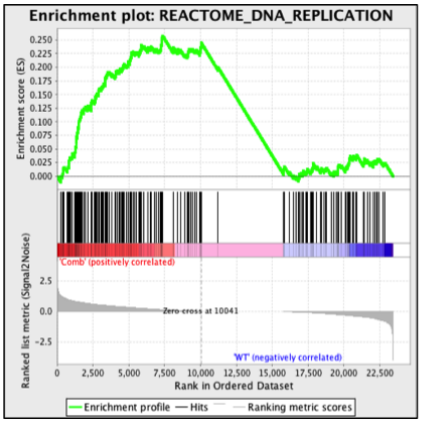
Figure 5.17: MkPs from combined mutant mice are positively enriched for immature megakaryocyte genes, but negatively enriched for genes associated with megakaryocyte/platelet maturity. GSEA plot, associated statistics and lists of genes driving positive and negative enrichment. In the heatmaps on the right, each row represents a gene and each column represents a biological replicate. Red = upregulated. Blue = downregulated.



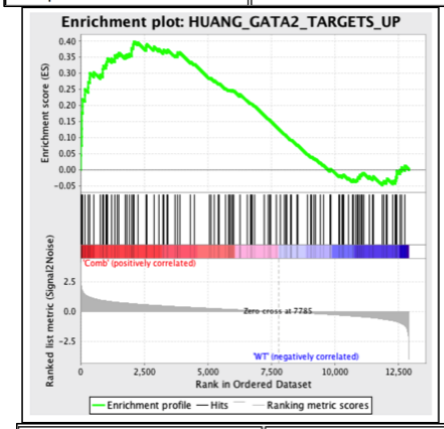
GeneSet	REACTOME_CELL_CYCLE
Enrichment Score (ES)	0.26517442
Normalized Enrichment Score (NES)	5.817118
Nominal p-value	0.0
FDR q-value	0.0



GeneSet	HALLMARK_E2F_TARGETS
Enrichment Score (ES)	0.36999404
Normalized Enrichment Score (NES)	5.933388
Nominal p-value	0.0
FDR q-value	0.0



GeneSet	REACTOME_DNA_REPLICATION
Enrichment Score (ES)	0.25795203
Normalized Enrichment Score (NES)	4.08954
Nominal p-value	0.0
FDR q-value	0.0



GeneSet	HUANG_GATA2_TARGETS_UP
Enrichment Score (ES)	0.35884243
Normalized Enrichment Score (NES)	1.4176685
Nominal p-value	0.019758508
FDR q-value	0.019758508

Figure 5.18: Gene set enrichment analysis of gene expression signature in Combined mutant MkPs. GSEA plots and associated statistics shown. Positive enrichment (red) in the combined mutant vs WT for Cell cycle, DNA replication, E2F targets and GATA2 targets.

5.2.1.7 Analysis of the genes differentially expressed between Gata1s only and combined Gata1s/cohesin mutant mice

It is notable that the number of significantly differentially expressed genes between Gata and Comb mutants was very low when a stringent p-adjusted value of $<$ 0.05 was applied. These genes, together with the Log2Fold change, p-adjusted value and known function are summarised in Figure 5.19. It is possible that additional differentially expressed genes are contributing to the observed phenotype (possibly in a combinatorial fashion) but do not reach statistical significance in this pair-wise analysis. These data would support a model in which the observed MkP phenotype is mainly driven by transcriptional dysregulation resulting from mutant Gata1s with the additional cohesin gene mutations contributing (potentially multiple) small but significant additive effects.

Gene	log2FoldChange	padj	Function
Haao	-1.005599594	0.043798979	3-hydroxyanthranilate 3,4-dioxygenase
Rgcc	-0.997393581	0.018736257	regulator of cell cycle
Col1a2	-0.791705279	0.043798979	collagen, type I, alpha 2
Sdpr	-0.784279862	0.017161541	Serum deprivation response protein
Nrgn	-0.716435559	0.000215985	neurogranin
H2-DMa	-0.564084283	0.004611608	histocompatibility 2, class II, locus Dma
Aoah	0.773277864	0.011182276	acyloxyacyl hydrolase
Car12	0.882777162	0.017899167	carbonic anhydrase 12
Gpr171	1.097305464	0.002671647	G protein-coupled receptor 171
P2ry14	1.425502115	9.56E-05	purinergic receptor P2Y, G-protein coupled, 14

Figure 5.19: Genes differentially expressed between Combined mutant and Gata1s only mutant (P-adjusted value <0.05). Differential gene list identified using a Wald Test in DESEQ2 with p-adjusted value <0.05. Log2Foldchange is relative to Combined mutant (negative = down in Combined mutant)

5.2.2 ATAC sequencing of sorted megakaryocyte progenitors

Whilst informative, analysing gene expression alone provides limited information about the activity of regulatory elements lying in non-coding regions of the genome. I therefore used ATAC-Seq (assay for transposase-accessible chromatin with high-throughput sequencing) to assess for the presence of genome-wide differences in chromatin accessibility between wild type and mutant mice.

ATAC-Seq was first described in 2013 (Buenrostro et al. 2013), and utilises a hyperactive mutant Tn5 transposase to tagment and simultaneously insert sequencing adaptors into regions of open chromatin (Figure 5.20). Tagged DNA fragments are then PCR amplified and sequenced. Sequenced reads can be used to infer regions of increased accessibility and map nucleosomes and transcription-factor binding sites. In comparison to other techniques used to explore the epigenome (ChIP-Seq, DNAse-Seq, FAIRE-Seq), ATAC-Seq has been shown to provide good quality sequencing data even with relatively small amounts of starting material.

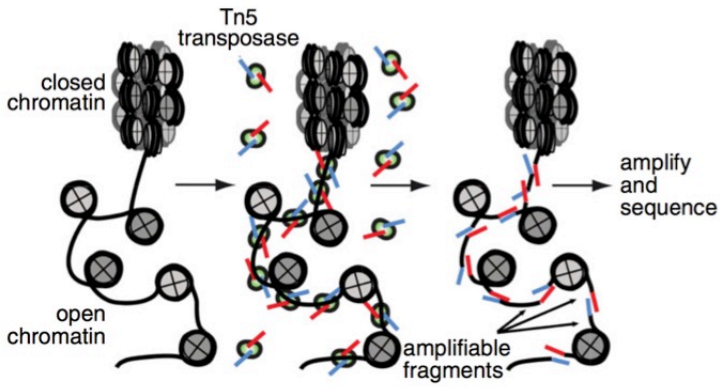


Figure 5.20: Assay for Transpose-Accessible Chromatin Sequencing (ATAC-Seq): Schematic. Hyperactive Tn5 mutant transposase cuts regions of open chromatin and simultaneously labels DNA fragments with sequencing adaptors for amplification and sequencing. Figure adapted Buenrostro et al. 2015with permission. Grey circles represent nucleosomes. Red and blue lines represent sequencing adaptors.

5.2.2.1 Optimisation and validation of protocol for low cell numbers

At the time of first performing this study, ATAC-seq protocols were established in the department for 50,000-100,000 cells but protocols for lower cell numbers were not in routine use. I therefore set out to optimise ATAC-seq for 1000-5000 cells, which is the approximate size of the MkP population in WT mice. Two of the major issues with scaling down the high cell number protocol (Buenrostro et al. 2013) are the risk of significant cell loss during pipetting steps and a high proportion of short mitochondrial reads taking up sequencing space and potentially biasing data. I therefore adapted the "FAST-ATAC" protocol described by Corces et al. (Corces et al. 2016) in which use of a gentle lysis agent minimises mitochondrial membrane breakdown and there are fewer pipetting steps. For lower cell numbers, I sorted cells directly into a reduced volume of lysis buffer in order to further limit cell loss (see summary Figure 5.21 and Materials and Methods).

	Method	Cell Type	Cell Number	Output	Reference
1	Lysis of pelleted cells with NP40 Transposition at 37°C PCR amplification and barcoding	50,000 K562	50,000, 10,000 5000 (poor results)	Tapestation Sequencing 78% mapping	Buenrostro et a. 2013 M. Suciu (Higgs lab, WIMM)
2	For 5000+ cell: Lysis (digitonin) and transposition of pelleted cells PCR amplification and barcoding 1000 cells: Sorted directly into reduced volume lysis/transposition mix	K562 Whole mouse BM Sorted mouse BM Human fetal liver AML	5000, 1000	Tapestation Sequencing 60-80% mapping for mouse BM samples	Corces et al. 2016

Figure 5.21: Optimisation of ATAC-sequencing protocol for low cell number.

The protocol was optimised for 1000 cells using mouse bone marrow and primary human AML samples. Libraries were quantitated using the D1000 high sensitivity Tapestation and sequenced. Data were trimmed and mapped to mm9, mm10 or

hg19 (Telenius, Consortium and Hughes, 2018). Mapping efficiency was 59-80% for all samples with 8-10% mitochondrial reads excluded from analysis. Representative Tapestation traces are shown for the ATAC libraries in Figure 5.22a-b, together with mapped sequencing data (Figure 5.23c). In general, there was a good correlation in peaks between published ENCODE DNase hypersensitivity data, 10,000 cell, 5000 cell and 1000 cell libraries. The protocol was further scaled down to 100 cells, however this was associated with a loss of accessibility data and not taken forward (see Figure 5.22c where some accessible regions are not detected in 100 cell data set).

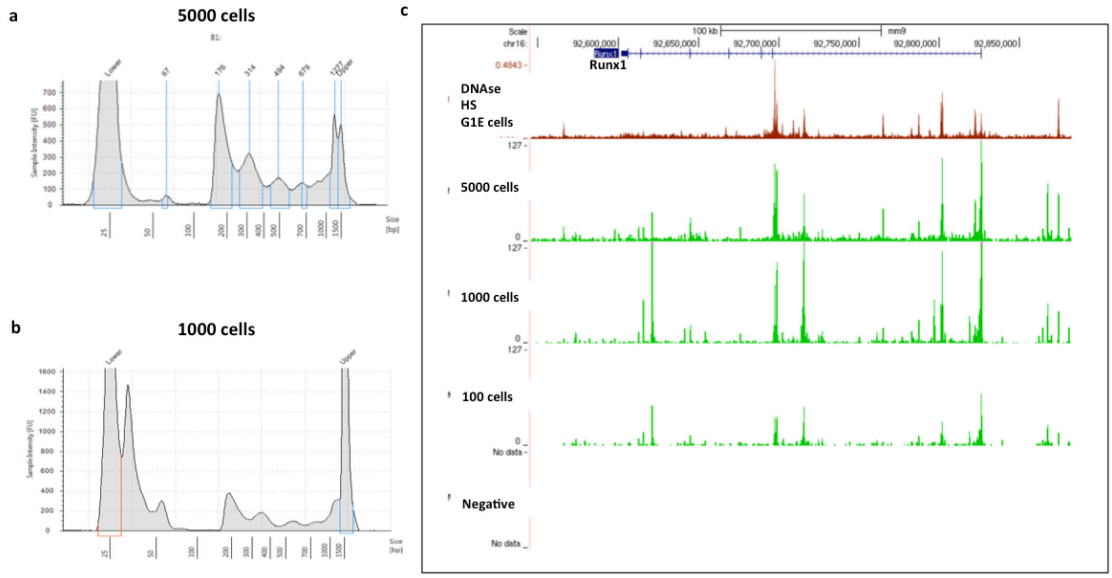


Figure 5.22: Validation of low cell number ATAC-sequencing protocol. Representative D1000 High sensitivity Tapestation Traces for 5000 mouse bone marrow cells (a) and 1000 mouse bone marrow cells (b) from ATAC-seq libraries showing appropriate size-distribution of fragments following tagmentation and amplification. Y axis reflects sample intensity from which library concentration can be calculated. X axis = fragment size (bp). Large peaks at the ends of axes represent markers. In an optimally tagmented sample, the majority of fragments will fall within a 100-300bp window representing the nucleosome-free region. The fragment size should then have clear periodicity, indicative of nucleosome occupancy. (c) Normalised ATAC-seq data for the Runx1 locus from 5000 cell, 1000 cell 100 cell and control libraries shown with downloaded published DNAse hypersensitivity data for G1E cell line from ENCODE.

5.2.2.2 Processing and quality control of ATAC-seq data

Using this optimised low cell number protocol, ATAC-seq was performed on sorted MkP cell populations sorted from WT (n=5), Rad (n=3), Gata (n=5) and Comb

mutants (n=4). Full details of sorting strategy, library preparation, sequencing and data analysis are outlined in Materials and methods. Filtered trimmed reads were mapped to the mm10 genome with an efficiency of 60-80% uniquely mapping reads. Adequate representation of all chromosomes was demonstrated by plotting the fraction of mapped reads per contig (Figure 5.23a). Interestingly, a high degree of variability in mitochondrial reads was seen despite the same lysis agent, method and volume being used for all samples. As the absolute number of mapped mitochondrial reads was low, this was assumed not to bias the overall dataset. Figure 5.23b shows the fragment size distribution across all samples. Consistent with the published standards for ATAC-seq data (https://www.encodeproject.org/atac-seq/) a large number of the reads were <100bp in length representing open chromatin and there was a dominant mononucleosome peak present in the fragment length distribution (spanning a single nucleosome so >147 bp and <2*147 bp) (Figure 5.23b). As seen in the Tapestation traces (figure 5.23), fragment size distribution showed clear periodicity, indicative of reads spanning nucleosome positions.

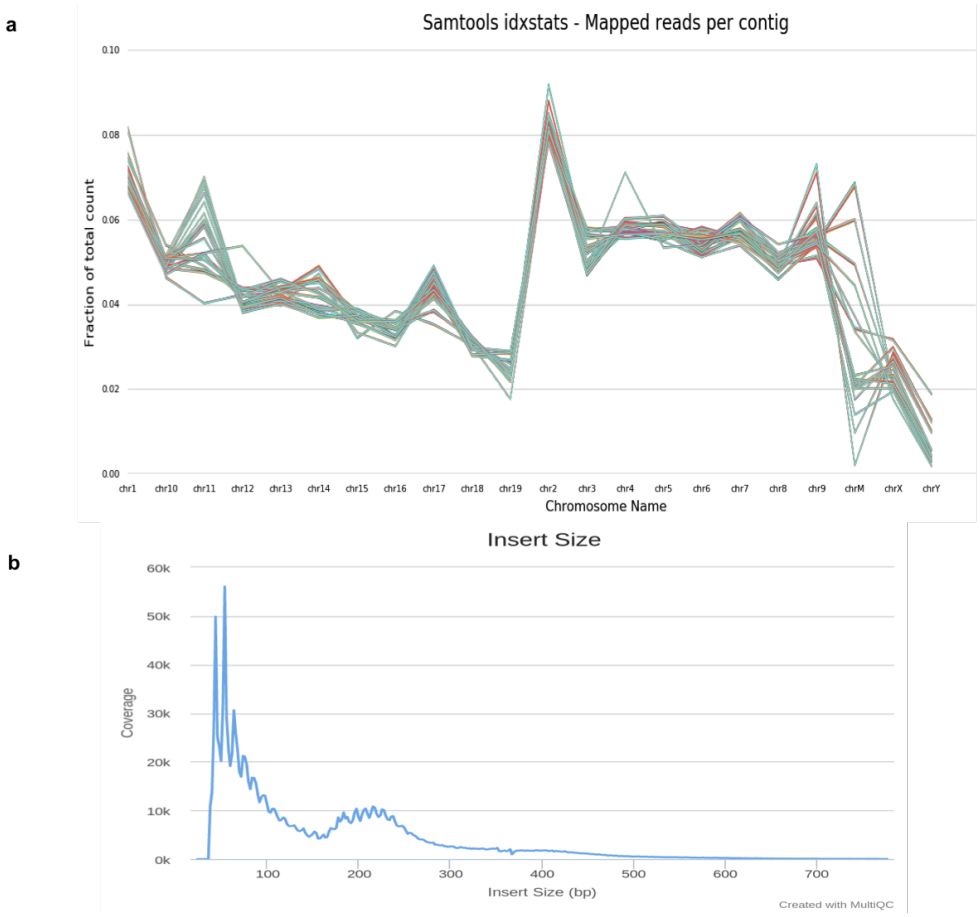


Figure 5.23: ATAC-sequencing of Megakaryocyte Progenitors: Coverage and Fragment size distribution. (a) Fraction of total read count mapping to each chromosome. Lines represent individual samples. (b) Fragment size distribution for combined ATAC-seq dataset. Coverage is mean read count per sample.

5.2.2.3 Identification of accessible sites: Peak calling

Mapped reads were filtered for duplicates, reads mapping to the mitochondrial genome and regions overlapping a known blacklist (Telenius, Consortium and Hughes, 2018).

To achieve the most robust peak set (in which a peak reflects pile up of sequencing reads and therefore an area of open chromatin), I first checked the consistency of replicates. Filtered bam files within each genotype group were highly correlated with each other by pairwise comparison (Figure 5.24), with a Pearson co-efficient of >0.96 for WT samples, >0.91 for Gata samples and >0.92 for Comb samples. Interestingly, Rad samples were less well correlated with a Pearson co-efficient of >0.81 - reflecting a higher degree of variability in distribution of mapped reads between biological replicates.

In the absence of obvious outliers, bam files from each genotype group were merged and peaks called on the merged files using Model Based Analysis of ChIP-Seq (MACS2) (Zhang et al., 2008). Total number of called peaks were WT = 83,003, Rad = 58,174, Gata = 86,456 and Comb = 86,811 with a stringent cut off of q=0.01 and WT = 94252, Rad = 67,342, Gata = 96,980 and Comb = 101,355 with q=0.05.

To ensure that peaks were being called appropriately, I performed a visual inspection of randomly selected peaks in individual sample replicates using the UCSC genome browser and also looked for correlation between identified peaks and areas of the genome expected to be both "open" (active histone marks, promoters, previously published ATAC-Seq data) and "closed" (repressive histone marks). As shown in Figure 5.25, peaks correlated well with areas predicted to have open chromatin and therefore be accessible to Tn5 tagmentation. They were also shown to be enriched at transcriptional start sites (TSS) (Figure 5.27, below).

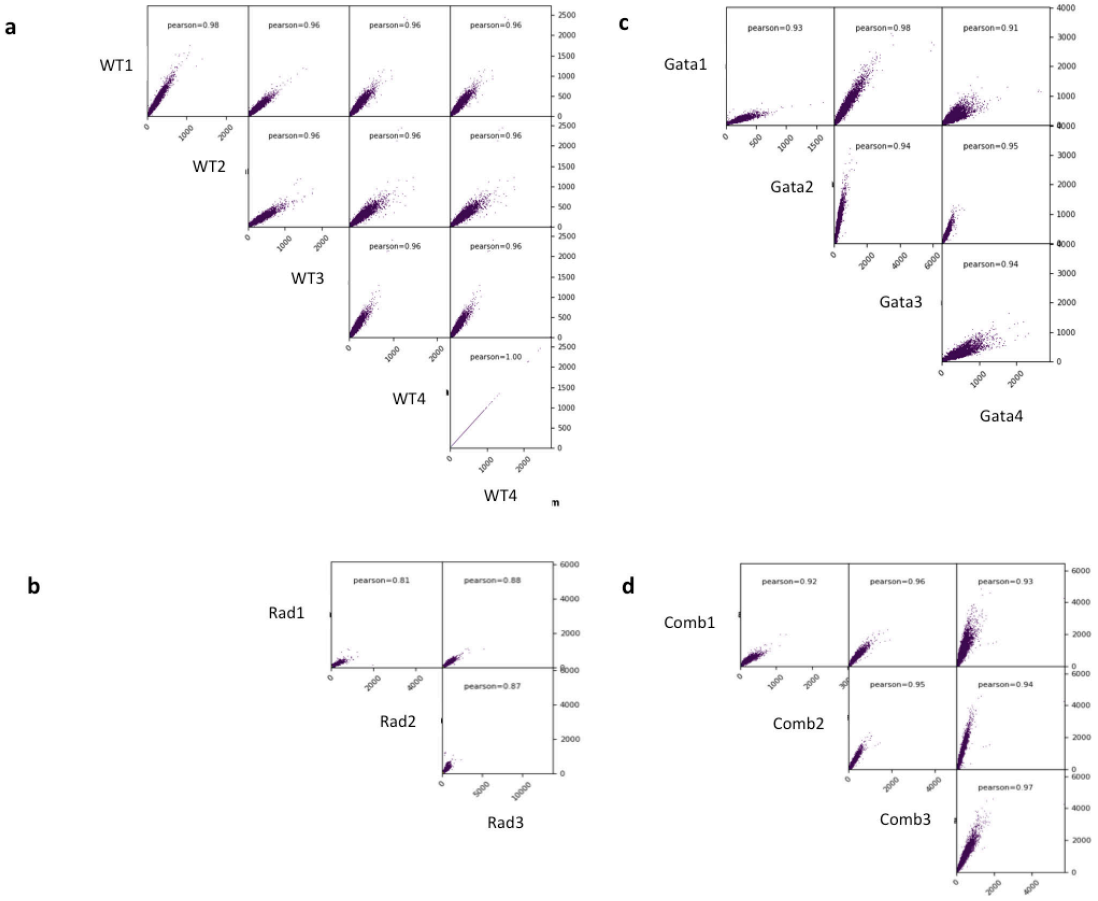


Figure 5.24: Pairwise correlation between biological ATAC-Seq replicates within each genotype group. Plots show pairwise comparison of mapped, filtered bam files for WT (a), Rad (b), Gata (c) and Comb (d) MkP samples. Dots represent reads within bioinformatically assigned bins. Pearson co-efficient for each pair is shown. Plot generated using deepTools.

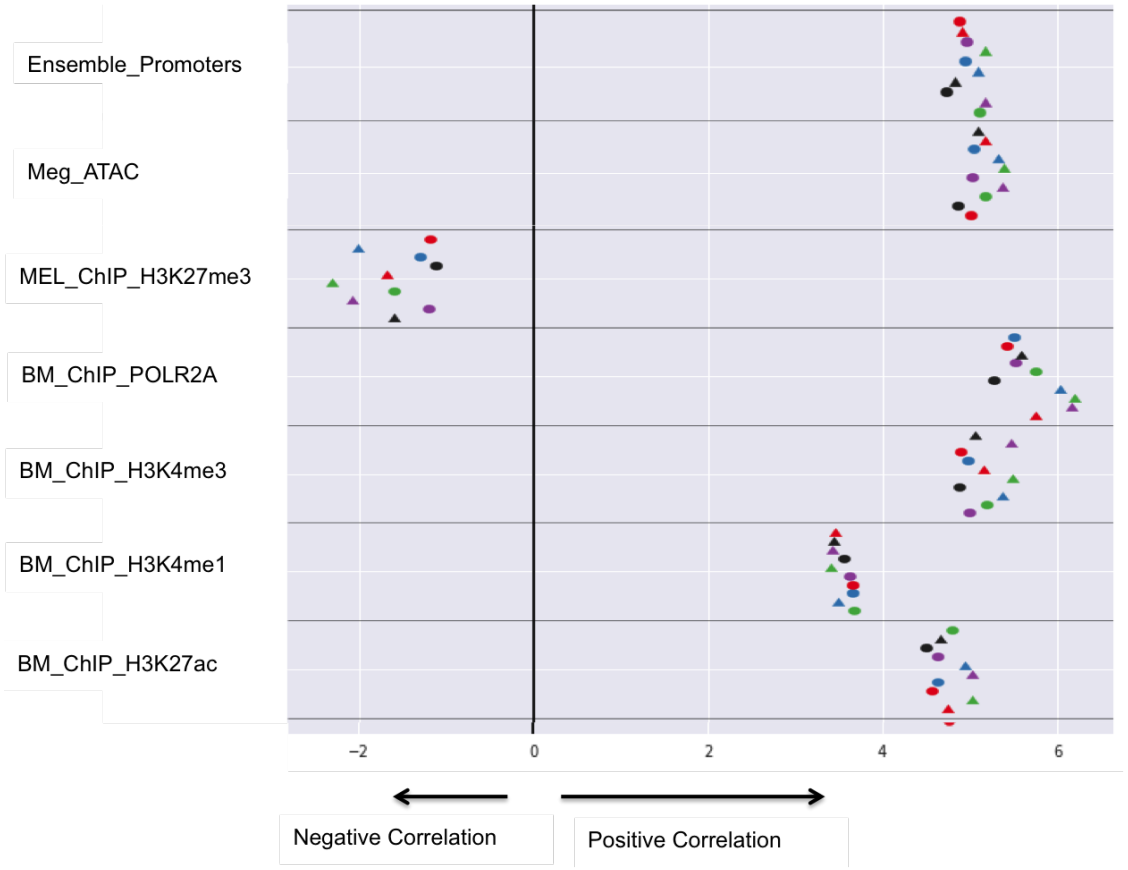


Figure 5.25: ATAC-seq reads from sorted MkP populations correlate with regions of accessible chromatin. Correlation of ATAC-seq peaks against downloaded datasets. Correlation score and direction shown on the X-axis. Each dot represents a merged ATAC-Seq library. Red = Comb, Green = Gata, Purple = Rad and Blue = WT. Circles q=0.01, Triangles q=0.05.

5.2.2.4 Haploinsufficiency of cohesin results in a global reduction in chromatin accessibility

It was notable that the total number of peaks called in the Rad21 mutant was lower than in WT, Gata and Comb MkP samples. This effect was apparent on visual inspection of merged and individual replicate tracks in the UCSC genome browser (Figure 5.26a), the fraction of reads in peaks (FRIP score) (Figure 5.26b) and when ATAC-Seq signal density was plotted against distance from transcription start sites (TSS) (Figure 5.27a). Equivalent results were obtained using alternative peak callers (data not shown). A similar effect of cohesin haploinsufficiency on global chromatin accessibility has been previously reported in a colon cancer cell line (Yan et al. 2013) and human CD34 enriched HSPCs (Mazumdar et al. 2015)), but not in MkP. Interestingly, a global decrease in accessibility was not observed in the Combined Gata1s/Rad21 mutant. Here the total number of called peaks, fraction of reads in peaks and global accessibility of TSSs was restored to that of WT and Gata1s only mutant samples.

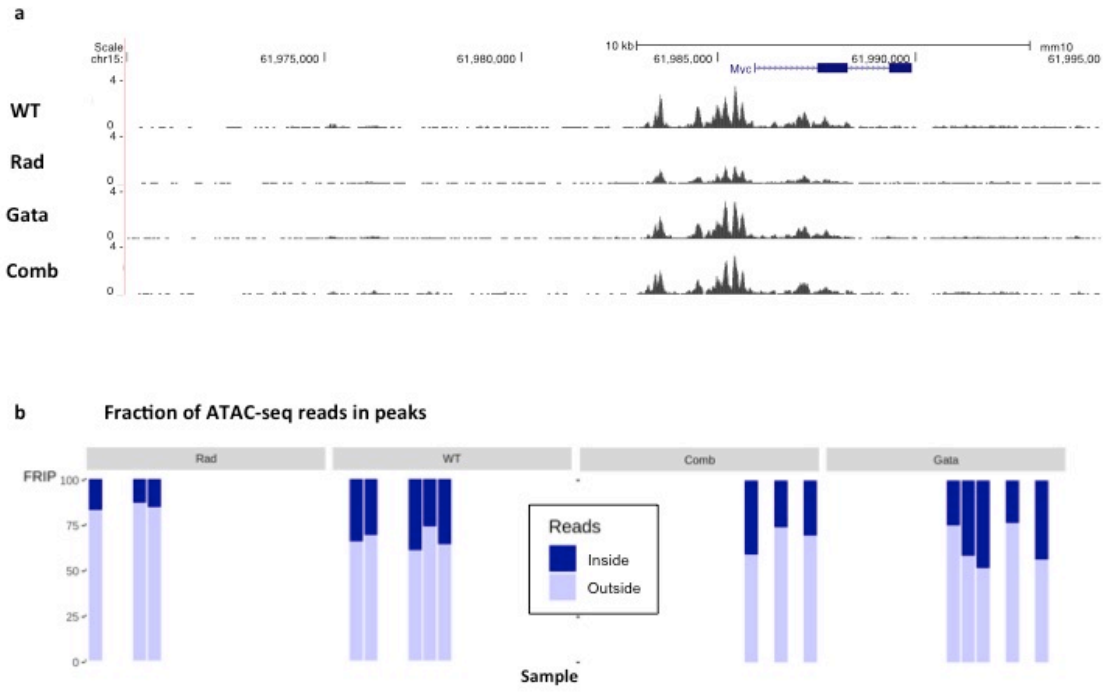


Figure 5.26: Haploinsufficiency of cohesin results in a global reduction in chromatin accessibility. (a) Representative track showing WT, Rad, Gata and Comb merged data normalised by library size for the myc locus. Reduced accessibility can be seen in the Rad track. (b) Summary histogram showing Fraction of ATAC-seq reads falling within called peaks. % of reads in peaks shown on Y axis. Each bar represents an individual sample. Sample groups shown in grey bar. <20% of Rad ATAC-seq reads fall in peaks, compared to >30% for all other samples.

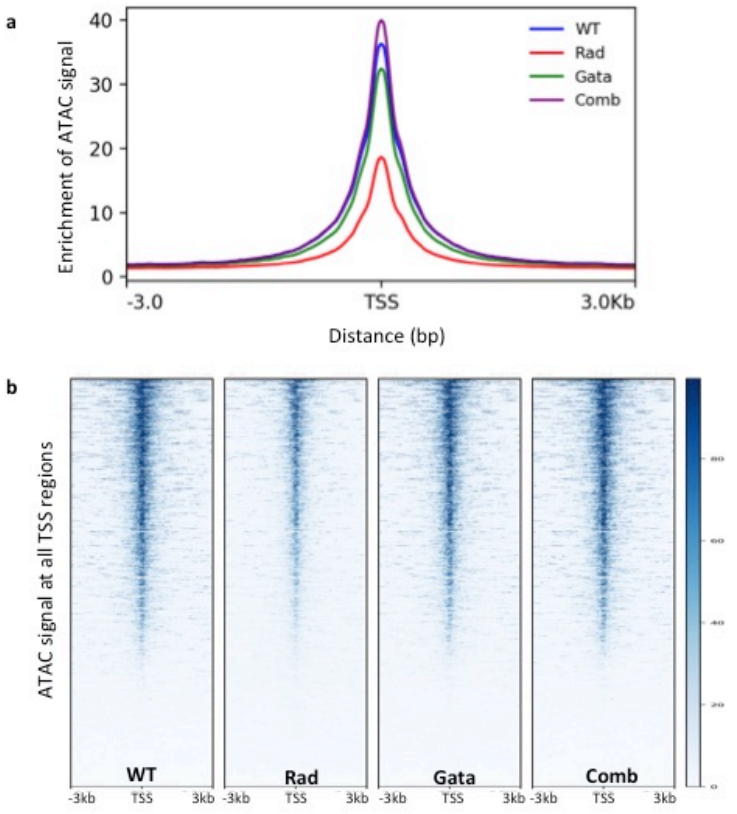


Figure 5.27: Genome wide chromatin accessibility at Transcriptional Start Site (TSS) regions. (a) Average diagram of genome-wide chromatin accessibilty at TSS regions in WT, Rad, Gata and Comb mutant MkP samples. (b) Heatmap of ATAC-seq signal at TSS regions.

5.2.2.5 Comb, Gata and Rad mutant samples can be segregated based on differences in chromatin accessibility

ATAC-Seq peaks were quantitated by counting fragments for each individual sample across a file containing all merged peaks to create a count matrix of accessible chromatin regions (Liao et al. 2014).

Principal component analysis (PCA) (Figure 5.28) showed that WT, Comb, Rad and Gata samples could be segregated based on chromatin accessibility, with the majority of the variance (PC1, 31%) separating WT from Cohesin mutant (Rad and Comb) samples. PC2 (27% of of the variance) appeared to separate Gata1fl (WT and Rad) from Gata1s mutant (Gata and Comb) samples, although the Gata cluster remained relatively poorly defined.

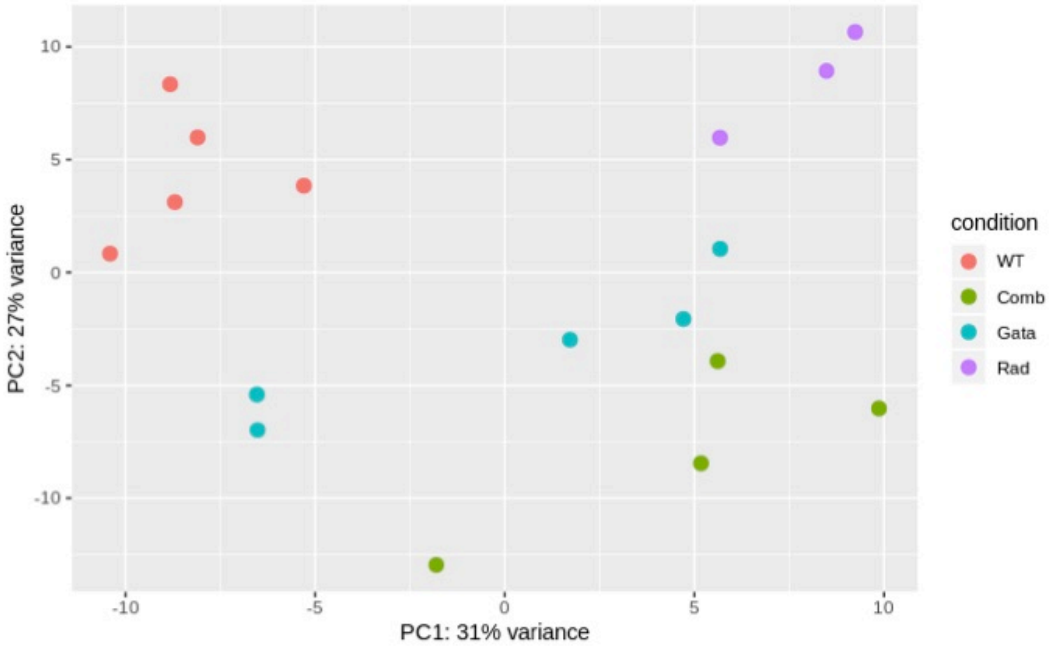


Figure 5.28: PCA of Megakaryocyte Progenitors using top 500 most variable ATAC-seq peaks. The percentage variance contributed by each principal component is indicated. Each dot represents a biological replicate.

5.2.2.6 Identification of differentially accessible sites

Differentially accessible sites were identified between control and mutant samples using Wald pairwise comparisons in DESeq2 (Love et al. 2014). The number of significantly (p-adjusted value <0.05) differentially accessible sites in each comparison is summarised in Figure 5.29. In keeping with the phenotypic and RNA-seq data previously discussed, the greatest difference in accessibility was noted between the Comb and WT, with only a small number of regions significantly differentially expressed between the Comb and Gata MkP groups.

Unlike RNA-seq data, there was no option to include a spike-in control, making accurate normalisation (particularly of the globally affected Rad mutant samples) challenging. A number of normalisation methods were trialled, however none produced consistent results with regards to the Rad group. Within the limitations of my analysis, Gata-WT, Comb-WT and Comb-Gata comparisons appeared to be fairly robust, regardless of the normalisation method used.

	Differentially Accessible Sites
Comb vs WT	3413
Gata vs WT	1475
Rad vs WT	1891
Comb vs Gata	63

Figure 5.29: Differentially Accessible Regions between genotypes. Summary table of number of significantly differentially accessible regions between genotypes (P-adjusted value <0.05).

For the purposes of visualisation only, merged samples were normalised by library size and also compared directly to un-normalised merged and individual tracks to ensure consistency. For differential accessibility analysis, both rlog (rld) and upper quartile normalisation were performed within DESEQ2. For the Gata-WT, Comb-WT and Comb-Gata comparisons, the differential accessibility lists were similar, regardless of normalisation method. For the Rad-WT and Comb-Rad comparisons, the same differentially accessible regions were called, however direction of accessibility (increased or decreased) was influenced by normalisation method, suggesting under or over compensation of the observed global change.

5.2.2.7 Gata1s and cohesin gene mutations together alter chromatin accessibility

In total, 3413 differentially accessible sites were identified between Comb and WT MkP samples (P-adjusted value <0.05). Of these, 2199 represented increased accessibility in the Combined mutant compared to wild type, and 1214 decreased accessibility. Representative examples of promoter regions with significantly increased and decreased accessibility in the combined mutant are shown in Figure 5.30, together with correlated gene expression data. Klf1, which demonstrates decreased expression and decreased promoter accessibility in the combined mutant is a key erythroid transcription factor and direct target of Gata1. Cpa3 (which shows increased expression in combined mutant cells and increased promoter accessibility) is a carboxypeptidase usually expressed in mast cells and another GATA target. Interestingly, expression is normally repressed by Gata1 and increased by Gata2 binding (Doré et al. 2011) suggesting that the observed increase in expression in the combined mutant may result from ineffective repression by Gata1s and increased Gata2 activity.

5.2.2.8 Annotation and enrichment analysis of ATAC-seq data

Appropriate annotation of ATAC-seq peaks is challenging, as accessible regions may represent cis-regulatory regions distant from their interacting gene and promoter. In my analysis, promoter regions (defined as 1Kb up or down stream of the TSS) were identified using the Homer annotatePeaks software (Heinz et al. 2010) and the ChIPpeakAnno package (Zhu et al. 2010).

Enrichment analysis of the lists of gene promoters showing increased accessibility in the Combined mutant compared to WT controls was performed using EnrichR (Chen et al. 2013). As shown in Figure 5.31, promoter sites with decreased accessibility in the Combined mutant compared to WT were enriched for Gata1, Gata2, CTCF, Rad21 (p=0.07) and Smc3 (p=0.07) binding based on Encode and ChEA datasets - perhaps supporting a hypothesis that genes displaying reduced promoter activity in the combined mutant are Gata and/or cohesin targets. By

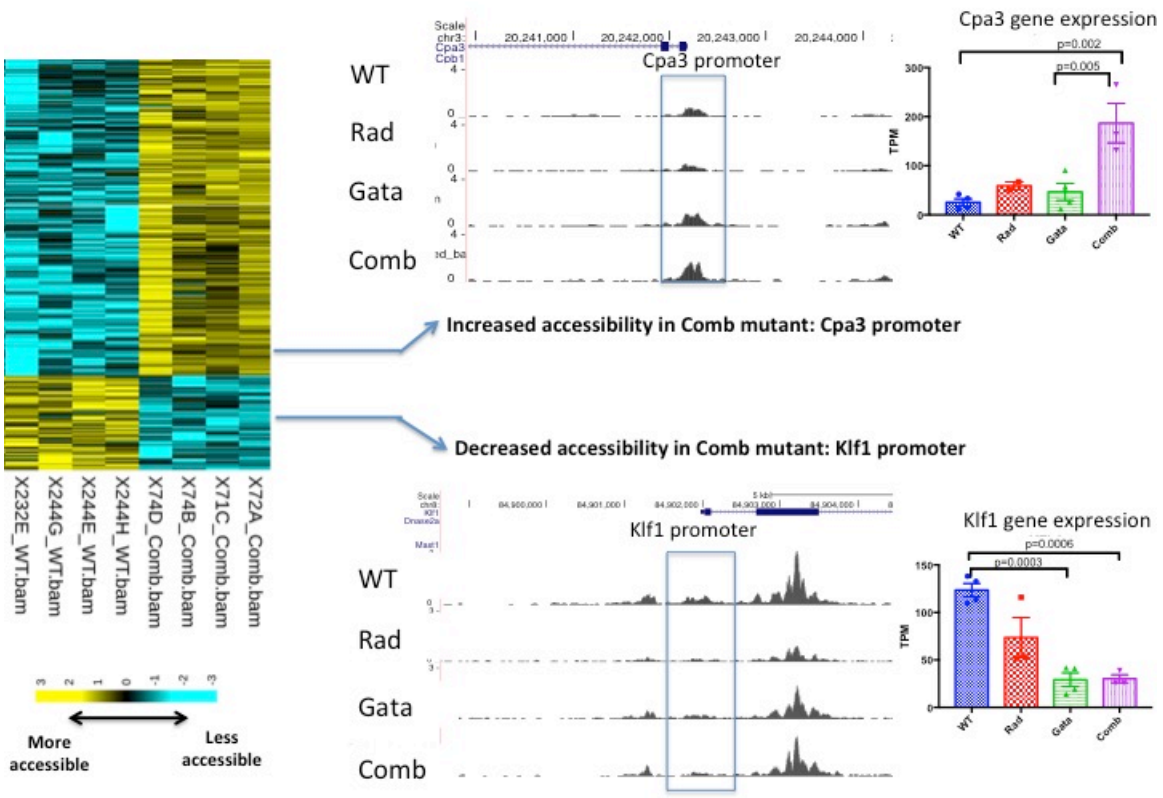


Figure 5.30: Gata1s and cohesin gene mutations together alter chromatin accessibility: heatmap and representative tracks. Heatmap of significantly differentially accessible regions identified by ATAC-sequencing in MkP of Combined mutant vs WT. Differential gene list identified using a Walt Test in DESEQ2 with p-adjusted value <0.05. Columns represent biological replicates of WT (left) and Comb (right) samples. Rows represent genes. Colour scale is normalised per row with scale shown below the heatmap. Representative UCSC genome browser tracks for regions with increased and decreased accessibility are shown on the right (Cpa3 and Klf1 promoters respectively). Tracks represent merged replicates for each genotype normalised by library size. Boxed region indicates the position of the promoter. Graphs on the right show gene expression of these genes calculated as transcripts per million (TPM) for WT, Rad, Gata and Comb mutants. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown).

contrast, promoter sites with increased accessibility in the Combined mutant showed enrichment for Lmo2, Runx1, Gata2, Lyl1, Meis1 and Gata1 ChIPseq data sets (Chea 2016) (Figure 5.32). Notably, promoters showing both increased and decreased accessibility were enriched for Gata2 binding - perhaps reflecting the positive and negative regulatory roles of this transcription factor.

ENCODE and ChEA Consensus TFs from ChIP-X

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	GATA1_CHEA	0.001903	0.1979	2.10	13.17
2	SMAD4_CHEA	0.002516	0.1308	2.29	13.73
3	STAT3_CHEA	0.003873	0.1343	3.53	19.61
4	ZC3H11A_ENCODE	0.01514	0.3936	3.46	14.50
5	GATA2_CHEA	0.02764	0.5749	1.73	6.23
6	SOX2_CHEA	0.02847	0.4934	1.73	6.15
7	CTCF_ENCODE	0.04460	0.6627	1.40	4.34
8	RELA_ENCODE	0.04665	0.6065	1.84	5.65
9	SMC3_ENCODE	0.07188	0.8307	1.44	3.78
10	RAD21_ENCODE	0.07563	0.7865	1.41	3.64

Figure 5.31: Promoter sites with decreased accessibility in the combined mutant are enriched for Gata and cohesin binding sites. Table of top 10 most enriched transcription factor pathways generated using the EnrichR package (Chen et al. 2013) based consensus binding from Encode and ChEA data sets.

ChEA 2016

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	LMO2_20887958_ChIP-Seq_HPC-7_Mouse	1.345e-8	0.000008675	2.54	46.02
2	RUNX1_20887958_ChIP-Seq_HPC-7_Mouse	7.601e-8	0.00002451	2.91	47.76
3	RUNX1_22412390_ChIP-Seq_EML_Mouse	0.000001649	0.0003544	2.16	28.73
4	GATA2_22383799_ChIP-Seq_G1ME_Mouse	0.000004183	0.0006746	2.11	26.07
5	GATA2_20887958_ChIP-Seq_HPC-7_Mouse	0.00002800	0.003612	2.10	21.97
6	MECOM_23826213_ChIP-Seq_KASUMI_Mouse	0.00007697	0.008274	1.94	18.40
7	LYL1_20887958_ChIP-Seq_HPC-7_Mouse	0.0008603	0.07927	2.38	16.80
8	GATA1_22383799_ChIP-Seq_G1ME_Mouse	0.001171	0.09443	1.74	11.72
9	MEIS1_20887958_ChIP-Seq_HPC-7_Mouse	0.001329	0.09523	1.88	12.48
10	SMRT_22465074_ChIP-Seq_MACROPHAGES_Mouse	0.004345	0.2803	1.63	8.87

Figure 5.32: Promoter sites with increased accessibility in the combined mutant are enriched for key transcription factor pathways. Table of top 10 most enriched transcription factor pathways generated using the EnrichR package (Chen et al. 2013) based consensus binding from ChEA data sets.

5.2.2.9 Integration with gene expression data

To assess how well this accessibility data correlated with the gene expression data previously described, I compared the total number of annotated "promoter" peaks to genes showing moderate - high expression (TPM >10) from each genotype group. The number of expressed genes with TPM >10 was very similar between groups (Figure 5.33). Number of annotated promoter ATAC peaks was similar for WT (15,794), Gata (15803) and Comb (15870) samples but reduced in the Rad mutant group (14541). For all genotypes (including Rad), the majority (87-88%) of moderate-high expressed genes were represented by an annotated ATAC peak suggesting good correlation between these datasets. The annotated ATAC-seq peaks which did not pair with an expressed gene may reflect regulatory regions located >1kb from their target promoter and therefore inappropriately assigned to the nearest gene. Additionally, changes in chromatin may precede changes in gene expression and therefore might be better represented in the upstream cell compartment (PreMeg-E). Finally, regions of open chromatin also only represent areas with the potential to bind trans-acting factors and activate gene expression. Correlation with ChIP-seq data would be needed to confirm whether or not this was the case.

I next compared the list of promoter annotated ATAC peaks found to be significantly different between Comb and WT samples (n=191) with the list of significantly differentially expressed genes (n=665). Only 12 genes (*Cst7, Rpp21, Cd34, Lrmp, Clcc1, Tubb2a, Alad, Clec4d, P2ry14, St7, Srm and Dhrs3*) were included in both lists. These genes are potentially of interest, including the immature HSPC marker CD34, genes involved in pathways regulating cell growth and differentiation (*Clcc1, P2ry14, Srm, Dhrs3*) and regulators of eosinophil differentiation (*Cst7* (Matthews et al., 2016)) and haem metabolism (*Alad). Tubb2a* encodes a tubulin protein - important for the terminal megakaryopoiesis and formation of the platelet cytoskeleton and St7 is a tumour suppressor gene, implicated in some cases of ovarian and liver cancer. However, it seems unlikely that this short list is a true representation of the genes and pathways most relevant to the observed Gata1s/Rad21 phenotype. As discussed above, this may reflect

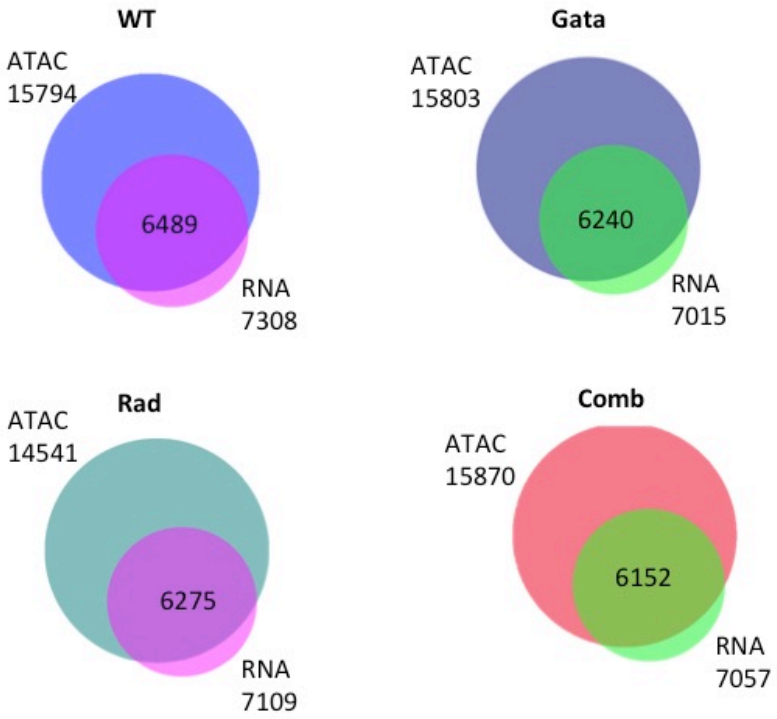


Figure 5.33: Overlap between RNA-seq and promoter annotate ATAC-seq datasets. Venn diagrams showing overlap between promoter annotated ATAC-seq regions and moderate-highly genes expressed from RNA-seq data from each genotype group. ATAC-seq peaks annotated using Homer. RNA-seq lists defined as genes with a normalised TPM of >10. Numbers outside the circles represent total number of genes. Numbers inside circles represent overlap.

the limitations of ATAC-seq peak annotation, the stringency of differential analysis and also the omission of non-promoter regulatory elements from peak annotation and analysis.

5.2.2.10 Motif analysis

In order to explore the accessible chromatin sites lying outside of promoter regions, I first performed transcription factor motif analysis on peaks with significantly increased and decreased differential accessibility in combined mutant cells compared to WT, using MkP ATAC peaks detected across all samples as a background dataset. In the absence of a robust list of murine MkP enhancers, I also attempted to enrich for distant cis-regulatory elements (enhancers) by re-running the analysis to include only peaks falling >1000bp from TSSs.

Statistically significant enrichment for the key haematopoietic regulators Runx1, ETS1, ERG, Meis1, Fli1, ETV1 and PU.1 was noted when all peaks were analysed (Figure 5.34a), suggesting that the combination of Gata1s and cohesin haploinsufficiency may result in perturbation of multiple transcription factor pathways. This is broadly consistent with the enrichment analysis described above for RNA-seq data and promoter sites showing increased accessibility in ATAC-seq data from Comb vs WT mice.

The data described so far also supports up-regulation of the Gata2 pathway in Comb mutant cells - based on increased expression of Gata2 itself (Figures 5.5 and 5.15), up-regulation of Gata2 targets in RNA-seq data (Figures 5.16 and 5.18) and known promoters falling within accessible regions in ATAC-seq data (Figure 5.32). This up-regulation appears to be greater than caused by Gata1s alone suggesting a co-operative role of cohesin. Interestingly, therefore GATA2 binding motifs were significantly enriched, but only when analysis was focused on sites >1kB from the TSS (Figure 5.34b), perhaps suggesting that aberrant Gata2 activity may be perturbing MkP gene regulation predominantly via distal rather than proximal regulatory elements. Clearly additional experiments would be needed to test this hypothesis further.

a

Motif	Name	P-value	Target Sequences with Motif
	Runx1	1e-68	646
	ETS1	1e-22	687
	ERG	1e-22	874
	Meis1	1e-22	620
	Fli1	1e-29	668
	ETV1	1e-14	403
	PU-1	1e-18	764

b

Motif	Name	P-value	Target Sequences with Motif
	Runx1	1e-40	572
	Erg	1e-34	920
	Fli1	1e-34	700
	ETS1	1e-23	697
	Gata4	1e-19	441
	Gata2	1e-18	345
	Gata3	1e-16	547
	ETV1	1e-15	782
	Gata1	1e-14	312

Figure 5.34: Chromatin regions which gain accessibility in combined mutant cells are enriched for key transcription factor binding motifs. (a) Motif enrichment search performed using all sites significantly differentially accessible between Comb and WT MkP. (b) Motif enrichment search performed using sites significantly differentially accessible between Comb and WT MkP and located >1kB from transcriptional start sites, enriching for distal regulatory elements. Motif enrichment search performed in Homer. p-value reflects enrichment above a background of all MkP ATAC-seq peaks.

5.2.2.11 Integration with Gata1/Gata1s ChIP-seq data

It has been postulated that differential Gata1 vs Gata1s binding and consequent transcriptional activation/repression may (at least in part) underlie the Gata1s phenotype (Chlon et al. 2015, Ling et al. 2019). Due to the low number of cells in the MkP compartment, generating a ChIP-seq data set from WT, Rad, Gata and Comb mutant MkP has proved challenging. In the first instance therefore, I compared my chromatin accessibility data with published Gata1/Gata1s ChIP-seq data from G1ME cells - a Gata1-null cell line which shows both erythroid and megakaryocytic differentiation potential (Chlon et al. 2015). Analysis of of data downloaded from the GEO archive (accession number: GSE64327), revealed a similar base-wise overlap of Gata and Gata1s ChIP-seq peaks to that reported by the authors, with the majority of Gata1s peaks also appearing in the Gata1FL dataset (Figure 5.35).

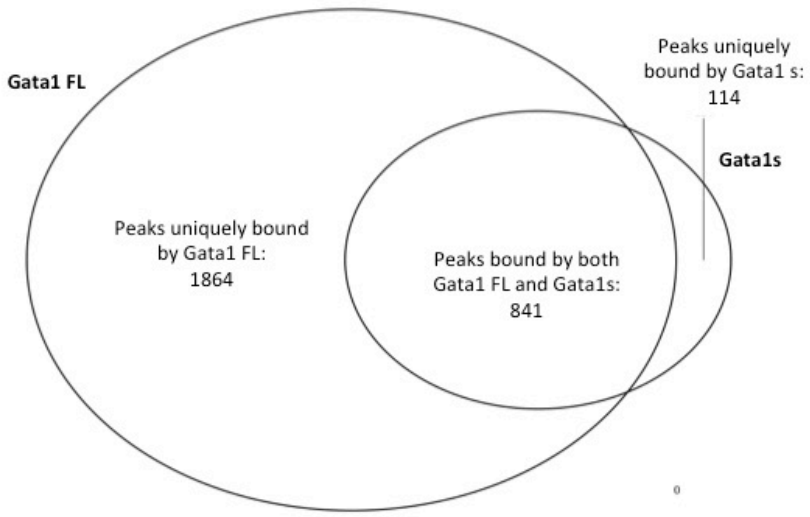


Figure 5.35: Overlap of GATA1 and GATA1s ChIP-Seq peaks from G1ME cells. Venn diagram represents base-pair overlap between ChIP-seq peaks from Gata1 FL and Gata1s expressing G1ME cell lines. Numbers of peaks are as shown. Data from Chlon et al. 2015.

I next intersected these peaks with with my own ATAC-seq dataset. In total, 105 of the 1864 ChIP-seq sites shown to lose Gata1s binding (uniquely bound by Gata1 FL) also showed decreased chromatin accessibility in the combined mutant compared to the WT. Although absolute sequence numbers were low, these peaks were significantly enriched for GATA motifs (Figure 5.36a) implying that loss of the Gata1 N-terminus may abrogate GATA binding - at least at a subset of sites. Interestingly, 62 of 114 the sites which gained Gata1s binding (uniquely bound by Gata1s) also showed increased accessibility in the combined mutant compared to the WT. These were enriched for alternative transcription factor motifs (NF-E2, Pbx) (Figure 5.36b) perhaps supporting a co-regulatory role for additional transcription factors in the Gata1s/cohesin mutant MkP.

a

Motif	Name	P-value	Target Sequences with Motif
	Gata3	1e-11	37
	Gata2	1e-8	26
	Gata1	1e-7	29
	Gata4	1e-7	27
	Gata:SCL	1e-5	11

b

Motif	Name	P-value	Target Sequences with Motif
	NF-E2	1e-5	27
	GATA:SCL	1e-5	10
	Gata3	1e-3	23
	Pbx3	1e-3	10

Figure 5.36: Transcription factor bindings motifs associated with Gata1/Gata1s binding and chromatin accessibility. (a) Motifs enriched at sites showing both decreased chromatin accessibility in the combined mutant compared to WT and loss of occupancy in Gata1s. (b) Motifs enriched at sites showing both increased chromatin accessibility and gain of occupancy in Gata1s. Motif enrichment search performed in Homer. p-value reflects enrichment above a background of all MkP ATAC-seq peaks.

This analysis also supported a role for distant regulatory regions driving the Combined Gata1s/Rad21 mutant phenotype. When mapped to annotated regions of the genome, ChIP-seq/ATAC-seq overlap sites were almost exclusively located in intronic and intergenic regions, with <10% falling at known promoter sites (Figure 5.37b). By contrast, around 20% of ATAC-seq peaks which did not overlap with differentially bound Gata1/Gata1s sites could be annotated to promoters and promoter proximal elements (Figure 5.37a).

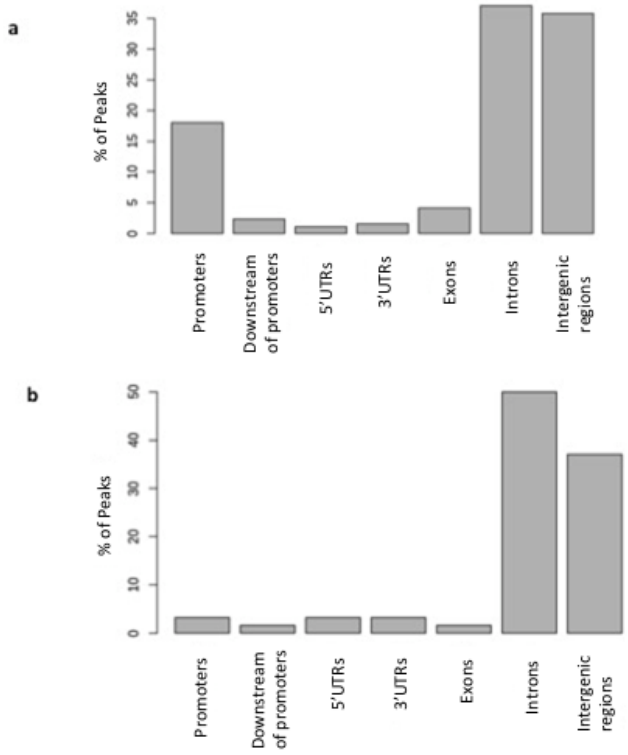


Figure 5.37: ATAC-seq Peaks differentially expressed between Comb and WT mutant MkP which overlap with regions of differential Gata1/Gata1s binding are mainly located in intronic and intergenic regions. Graphs showing the percentage of peaks mapped to promoters, downstream promoter elements, 5'UTR, 3'UTR, Exons, Introns and Intergenic regions for (a) all Gata1 and Gata1s ChIP-seq peak regions and (b) Peaks showing differential Gata1/Gata1s binding which overlap with significantly differentially accessible regions in ATAC-seq data from Combined mutants compared to WT. Y axis represent % of reads mapped to each category. Scale is 0-35 in (a) and 0-50 in (b). Histogram compiled using ChIPpeakAnno software (Zhu et al., 2010).

As a final step, these integrated data were visualised in the UCSC genome browser. This analysis will be discussed in further detail in the next chapter when

considering potential mechanisms of co-operation at specific gene loci.

5.2.3 Discussion

Taken together, these RNA-sequencing and chromatin accessibility data support a co-operative role for Gata1s and haploinsufficiency of Rad21 in altering gene expression profiles and expanding the MkP population.

Specifically, the gene expression signature of MkP cells from combined Gata1s/Rad21 mice reflects the abnormal proliferation and differentiation observed in the expanded MkP population and suggests perturbation of key transcription factor pathways. Both unbiased PCA analysis and interrogation of significantly differentially expressed genes indicates that Gata1s plays an important role in this phenotype, negatively enriching for erythroid genes and pushing towards proliferative megakaryopoiesis. By contrast, haploinsufficiency of cohesin appears to have multiple small but significant effects in the cell (some of which may be missed by stringent differential analysis criteria), co-operating with Gata1s to perturb gene expression over and above the impact of either mutation alone.

Using a protocol optimised for low cell numbers, it was possible to assess the global chromatin accessibility in Comb, Gata, Rad and WT MkP populations. Significantly, genotype groups could be segregated by unbiased PCA of ATAC peaks, and, as reported by others, cohesin haploinsufficiency resulted in global loss of accessibility, although this is the first time it has been documented in the MkP. Interestingly, this effect was not apparent in the combined Gata1s/Rad21 mutant - perhaps suggesting a dominant effect of Gata1s and/or an effect on chromosome topology resulting from cohesin depletion in this cellular context.

Given the likely genome wide effects of cohesin haploinsufficiency, normalisation of sequencing data to allow accurate comparisons between individual samples and sample groups is an important issue. For the RNA-sequencing analysis, I included an exogenous spike-in control - although this approach is subject to technical variation (Risso et al., 2014). For the ATAC-seq data, no such spike-in control is available. I trialled normalising using a number of different methods (library

size, upper quartile, rlog and RUVseq and using regions of the genome showing no change in gene expression) however quantitative results for the mutant Rad samples (where a global change in accessibility was noted) remained variable, and the presented data should be viewed in light of this caveat.

A number of key haematopoietic transcription factors were either up- or down-regulated in the combined mutant model, including Gata2 and the NURD complex components MTA2 and MTA3 (increased expression compared to WT) and Runx1 (decreased expression compared to WT). Enrichment analysis of the genes differentially expressed between the combined mutant and wild type demonstrated enrichment for some of these transcription factor pathways (including Gata2, Runx1 and E2F targets). ATAC-seq data supported these findings, with enrichment for Runx1, Gata2, Meis 1 and ETS1 transcription factor binding sites in regions showing differential accessibility in the combined mutant. It would seem reasonable to assume that dysregulation of these highly relevant transcription factor pathways is contributing to the observed phenotype. As many of these factors (directly or indirectly) regulate each other, understanding the exact mechanism is likely to be complex.

Integration with published Gata1/Gata1s ChIP data showed that many of the significantly differentially expressed genes between combined and WT MkP cells were Gata1 targets, suggesting that differential function of Gata1s at these sites might be contributing to the observed phenotype, perhaps through dysregulation of transcription factor pathways such as Gata2 and Runx1. These factors may be acting at regulatory loci distant to mapped promoters and additional topological studies would be useful to explore this further. Either way, the additive effect of cohesin loss is likely to be complex, with multiple (potentially subtle) effects throughout the genome driving further gene dysregulation - potentially in a locus-specific manner.

One approach to moving closer to understanding the mechanism of co-operation would be to identify a key locus (or loci) where Gata1s and haploinsufficiency of Rad21 together perturb gene expression in a synergistic (or additive manner) and try to understand the (possibly locus-specific) mechanism of interaction. To be

relevant, these loci should encode a functionally important gene that explains, at least in part, the observed MkP expansion. From my dataset, good candidates are the dysregulated transcription factors Gata2, Runx1 and MTA3. An approach to studying these loci in more detail will be discussed in the next chapter as part of ongoing and future work.

6
Discussion

Contents

The mechanisms by which mutations in haematopoietic stem and progenitor cells may act together to perturb pathways of proliferation and differentiation remains an important question in leukaemia biology. This thesis explores the co-operation between two mutations enriched in Myeloid Leukaemia of Down Syndrome: Gata1s and haploinsufficiency of the cohesin complex component Rad21.

6.1 ML-DS is enriched in loss of function cohesin gene mutations

My first aim was to characterise the type and spectrum of cohesin gene mutations co-occurring with *GATA1s* in ML-DS patients. To do this I analysed a large cohort of TAM and ML-DS samples by next generation sequencing. As expected, GATA1s mutation(s) were present in all TAM and ML-DS samples. The majority

of ML-DS samples carried at least one additional somatic variant. These variants were predominantly in genes encoding signalling molecules, epigenetic regulators and components of the cohesin complex. Cohesin complex mutations were largely mutually exclusive. The most commonly mutated cohesin genes were *STAG2, RAD21, SMC1A* and the associated transcriptional regulator *CTCF*. For *STAG2* and *RAD21*, the majority of variants were nonsense, frameshift or splice-site and predicted to result in loss of cohesin function. In a number of cases, it would appear that T21, *GATA1s* and a single cohesin gene mutation were sufficient for transformation to leukaemia.

These findings correlate well with previously published studies (Yoshida et al. 2013, Nikolaev et al. 2013) and provide a valuable resource of both robust sequencing data and curated samples for ongoing and future work.

The results are clearly not exhaustive. Although a subset of samples underwent whole exome sequencing, novel variants lying outside the targeted amplicon panel may not have been covered. It is also possible that some potentially oncogenic low VAF variants, large scale insertions or deletions and fusion transcripts may have been missed. As sequencing was performed on bulk bone marrow or peripheral blood samples, it is not possible to determine the progenitor populations in which mutations are acquired and which mutations co-exist within an individual cell. Sequencing analysis of sorted progenitor population and/or single cell studies would potentially be a highly informative next step. Equally, integrated analysis of DNA sequencing data from these samples with gene expression and chromatin accessibility studies might provide further insights into mechanisms of dysregulation.

Recent follow-on work performed by our collaborators has provided valuable insight into the functional consequences of some of these variants (Labuhn et al. 2019). A multiplex *in vivo* loss of function screen was performed in a murine model closely resembling TAM to assess for leukaemogenicity of 22 of the reported ML-DS variants (Labuhn et al. 2019). In brief, CRISPR-Cas9 genome editing was used to express Gata1s in mouse fetal liver cells. These cells were then exposed to a pool of guides targeted to key genes identified in our study and transplanted into

recipient mice. Loss of function of 18 of the 22 genes led to development of an ML-DS-like leukaemia with short latency, suggesting a co-operative effect with Gata1s. The spectrum of co-operating mutations was fairly similar to that seen in human patients, with the notable exception of the cohesin variants which were under-represented in this disomic model. This observation, which is consistent with my own findings, will be discussed in more detail below.

6.2 Gata1s plus haploinsufficiency of cohesin co-operate to cause a novel haematopoietic phenotype

Having established that Gata1s and haploinsufficient cohesin gene mutations commonly co-occur in ML-DS, my next aim was to explore how these mutations (alone and in combination) perturb haematopoiesis. To do this, I used a transgenic mouse model system, comparing the haematopoietic phenotype resulting from (i) Gata1s; (ii) heterozygous loss of the core cohesin component Rad21; (iii) the combination of Gata1s plus a *Rad21* mutation.

On their own, both Gata1s and Rad21 mutations were shown to cause a subtle haematopoietic effect. Gata1s mutant mice demonstrated impaired terminal erythropoiesis and increased megakaryopoiesis. The bone marrow Pre-megE and MkP compartments were moderately expanded and there was an increase in hyperlobulated megakaryocytes on bone marrow trephine. Although this is in keeping with the fetal phenotype (Li et al. 2005), it has not been previously reported in adult mice. Rad21 haploinsufficient mice displayed a decrease in the long term HSC and LMPP compartments. Although the absolute size of the MkP compartment was not increased, sorted MkP cells were more proliferative in culture and showed a trend towards decreased ploidy.

Significantly, combined Gata1s/Rad21 mutant mice displayed a more marked phenotype than either Gata1s or Rad21 mutants alone, supporting a role for mutational co-operativity. This was most clearly seen in the megakaryocyte progenitor population, which was significantly expanded in the combined mutant

compared to all other groups and displayed abnormal proliferation and differentiation *in vitro*.

It is notable that despite expansion and dysregulation of a relevant cell population, combined Gata1/Rad21 mutant mice did not develop leukaemia, even when aged to >12 months. This is consistent with the findings of Labuhn et al. (2019) and suggests that although Gata1s and cohesin mutations may perturb haematopoiesis in a co-operative manner, they are not sufficient for ML-DS in the absence of T21.

There are a number of possible explanations as to why trisomic cells might be particularly sensitive to reductions in cohesin. The extra dosage of trisomic genes on Hsa21 may influence cell proliferation and differentiation directly due to overexpression of Hsa21-encoded transcription factors such as RUNX1, ERG and ETS2, which have been implicated in haematological malignancies (Korbel et al. 2009). Cohesin may help to mediate this - through DNA-looping and/or maintenance of transcriptional domains. In the absence of a full dose of cohesin, this regulatory effect could be lost. Indeed, both my data (Figure 5.35) and published studies in cohesin mutant cells (Mazumdar et al. 2015) demonstrate increased binding site accessibility of Runx, ETS and ERG motifs in the context of cohesin haploinsufficiency, lending support to this hypothesis.

There is, however, evidence that the effects of Hsa21 may be more complex than a simple dosage effect. Letourneau et al. demonstrated global transcriptional differences resulting from T21, potentially affecting any or all of the other disomic chromosomes (Letourneau et al. 2014). This remains poorly understood and may result from many factors, including chromatin accessibility, methylation changes (Malinge et al. 2013) and/or the microenvironment and lead to dysregulation of multiple genes. Cohesin depletion may accentuate this effect - either by further disrupting chromatin structure and accessibility (Figure 5.28) or through additional multiple small changes in gene expression, which might cumulatively have a significant effect on the function of the cell.

Understanding these interactions is important but challenging. Trisomy was not included in the mouse model presented in this thesis for two reasons: Firstly, although a number of elegant mouse models of T21 have been created (discussed in Chapter 1.4.3), none fully recapitulate the human haematopoietic phenotype, potentially limiting the validity of results (Alford et al., 2010). Secondly, my aim was to explore the molecular co-operativity between Gata1s and cohesin gene mutations. The additional role of T21, whilst highly relevant, clearly adds an additional layer of complexity. Going forward, it will be important to address the role of Trisomy in ML-DS pathogenesis. This may be more effectively done using GATA1s trisomic human cells, with and without haploinsufficient cohesin gene mutations.

Other interesting questions concern the role of the cellular environment and the timing of *Gata1* and cohesin gene mutations. The work presented here was performed in adult mouse bone marrow progenitors, constitutively expressing Gata1s and conditionally depleted of Rad21 from E13.5 in VavCre-expressing haematopoietic and endothelial cells (Gan et al. 2010). Although not fully characterised, we do know that in ML-DS, *Gata1s* is acquired in fetal life, prior to the cohesin gene mutation. ML-DS is apparent in the post natal environment, and so this is a arguably a relevant model system, if not a perfect one. It would however be interesting to see whether the effects of the mutations are different if studied in fetal liver or yolk sac cells, and this work is currently underway. Another approach would be to study the timing of mutation acquisition, using inducible Cre-lox recombinases to conditionally deplete cohesin in Gata1s mice at different stages of development.

6.3 Gata1s and heterozygous loss of Rad21 alter gene expression profiles and chromatin accessibility in megakaryocyte progenitor cells

To address the question of mutational co-operativity, my final aim was explore how *Gata1s* and a haploinsufficient *Rad21* gene mutation might interact at a molecular level to cause expansion of the MkP compartment and abnormal proliferation and differentiation of cells within it.

Through integrated analysis of RNA-seq and ATAC-seq data from Gata1s, Rad21 mutant, Combined Gata1s/Rad21 mutant and WT MkP populations, I established that the combined effects of Gata1s and cohesin haploinsufficiency on both gene expression and chromatin accessibility were significantly greater than the effect of either mutation alone.

The gene expression signature of combined Gata1s/Rad21 mutant MkP was consistent with the observed phenotype - reflecting upregulation of genes relevant to cell cycling and early megakaryopoiesis and down regulation of genes implicated in terminal differentiation of megakaryocytes and erythroid cells. There was evidence of dysregulation of key transcription factors, including *Klf1*, *Runx1*, the NURD complex components *MTA2* and *MTA3* and *Gata2* (discussed in detail below) and genes differentially expressed between combined mutant and WT cells were significantly enriched for some of these transcription factor targets.

Analysis of chromatin accessibility data was challenging. The global decrease in accessibility observed in *Rad21* mutant cells meant that accurate normalisation of data was difficult to assess. Furthermore, annotation of accessible peaks was limited by the fact that the majority of peaks fell outside promoter proximal regions and a robust list of murine MkP enhancer elements was lacking. Despite these issues, it was possible to draw some interesting conclusions.

The global decrease in ATAC-seq signal seen in Rad21 mutant cells has been observed previously (Mazumdar et al. 2015, Yan et al. 2013) and is presumed to result from altered TAD and sub-TAD architecture. My own data (Chapter 5) and other studies (Viny et al. 2015, Mazumdar et al. 2015, Mullenders et al. 2015, Viny et al. 2019, Despang et al. 2019) suggests that this has only a relatively subtle effect on gene expression. Interestingly, in the combined Gata1s/Rad21 mutant, chromatin accessibility appeared restored to WT levels. The mechanism for this is unclear, but it could reflect an interesting interaction between Gata1s and cohesin. Chromosome topological studies (discussed below) would be potentially very interesting in this context.

For all genotypes, promoter proximal elements represented only a small fraction of the annotated ATAC-seq accessible regions, highlighting the importance of understanding distant regulatory elements and 3-dimensional genome structure. Analysis of annotated promoter sites showing differential accessibility in the combined mutant did, however, reveal: (a) Regions with reduced accessibility in the combined mutant were enriched for Gata1 and cohesin binding sites - suggesting that decreased Gata1s binding together with a reduced dosage of cohesin may result in decreased promoter activity at some loci; (b) Regions with increased accessibility in the combined mutant were enriched for binding of a number of key haematopoietic transcription factors known to promote megakaryopoiesis, including Runx1 and Gata2. To explore the large number of ATAC-seq peaks distant from promoter elements, I performed transcription motif analysis on differentially accessible sites. Again, enrichment was shown in the combined mutant for binding sites of key haematopoietic regulators, including Runx1, ETS, ERG and Gata2.

Finally, I integrated my sequencing data with published Gata1/Gata1s ChIP-seq data (Chlon et al. 2015). Interestingly, at sites where differential Gata1 vs Gata1s binding could be demonstrated, almost all ATAC-seq peaks were located in regions distant from transcriptional start sites, perhaps reflecting regulatory elements which would be more likely to be influenced by topological changes resulting from cohesin depletion.

Notably, both RNA-seq and ATAC-seq data revealed only subtle differences between the Combined mutant and Gata1s only mutant MkP. In part this may reflect a dominant effect of Gata1s in driving the observed phenotype. It may also suggest a role for multiple small effects caused by cohesin depletion which do not necessarily reach significance on statistical analysis.

Indeed, one hypothesis is that the observed MkP expansion may result from disruptions in lineage commitment - not just reflecting an expansion of cells adopting a megakaryocytic fate but rather loss of the ability to suppress other options. This is supported by the observations that: (i) when cultured *in vitro*, a proportion of combined mutant MkP cells express the erythroid marker TER119 - suggesting the

potential for erythroid differentiation; (ii) the mast cell gene *Cpa3* is upregulated in the combined mutant (shown by both RNAseq and ATACseq promoter accessibility) and (iii) there is evidence of enrichment for PU.1 binding in combined mutant MkP cells, which positively regulates Gata1 expression in mast cells (Takemoto et al. 2010). This is likely to be mediated by transcription factor pathways (e.g. Gata1/FOG-1/NuRD) perturbed in the combined mutant and it is tempting to speculate that (as recently demonstrated in cells undergoing erythroid differentiation), cohesin may play a role in the dynamic regulation of this process (Sasca et al. 2019).

A number of further experiments would help with the interpretation of these findings. In the above analyses, I have used published ChIP-seq data from a relevant cell line to augment my analysis. Ideally, I would obtain ChIP-seq data from sorted MkP - specifically to look for Gata1/Gata1s and CTCF/Rad21 binding and histone marks to enable identification of potential enhancer elements. To date this has been limited by cell number, however I am currently optimising a low cell number protocol to use on a larger bone marrow progenitor population upstream of MkP. Information on 3D chromosome topology would also be of great benefit - both to assess the global effect of cohesin on TAD boundaries in the context of Gata1s and also to examine interactions at specific loci. Current protocols for 3C and Capture-C require at least 100,000 cells and are therefore not practical in primary MkP populations. It may, however, be possible use newer protocols, optimised for low cell number, when these become available (J. Davies, M. Oudelaar personal communication). There is growing evidence that cohesin may play a be particularly important role in dynamic cellular processes, including erythroid differentiation (Sasca et al. 2019) and response to inflammation (Cuartero et al. 2018). Studying the combined effect of Gata1s and cohesin haploinsufficiency during megakaryocyte differentiation or in response to stress would be of interest. Finally, additional information might be gained from analysis of chromatin accessibility in the progenitor population preceding the MkP, as chromatin changes may precede changes in gene expression. As there is growing evidence that the PreMegE is a mixed population, containing cells primed for megakaryocyte or erythroid fate

(Psaila et al. 2016) this would require careful consideration. Taken together, these data are suggestive of a mechanism by which Gata1s and haploinsufficiency of cohesin co-operate at a molecular level to disrupt regulation of genes involved in haematopoietic cell proliferation and differentiation causing expansion and aberrant function of the MkP. Combining my findings with published data, it seems likely that the effect of Gata1s is due, at least in part, to differential binding of target genes, while depletion of cohesin results in relatively subtle topological effects which may influence the interaction of regulatory elements with promoters at multiple loci.

6.4 Ongoing and future studies: investigation of target loci

In order to explore this hypothesis, and start to understand the complex mechanism of co-operation between Gata1s and Rad21 it will be necessary to explore individual loci in detail. I will therefore conclude by discussing a few transcription factor loci which represent interesting targets for future studies.

6.4.1 Gata2

Gata2, another member of the GATA family of transcription factors, is expressed in haematopoietic progenitors, including early megakaryocytes, erythroid cells and mast cells. *Gata2* transcription is repressed by Gata1, which assembles complexes on regions -3.9kb, -2.8kb and -1.8kb upstream of the *Gata2* haematopoietic (1S) promoter (Orlic et al. 1995, Grass et al. 2003, Martowicz et al. 2005). In the transcriptionally active state, these regions are occupied instead by Gata2, which autoregulates its own expression. As normal haematopoiesis proceeds, levels of Gata1 increase and levels of Gata2 decrease (Cantor et al. 2002). Additional regulatory elements have been identified at -77kb and +9.5kb (Figure 6.1). The +9.5 enhancer is important for HSC emergence from the vascular endothelium (Gao et al. 2013), while the -77kb enhancer is important for myeloid and megakaryocyte/erythroid differentiation (Johnson et al. 2015).

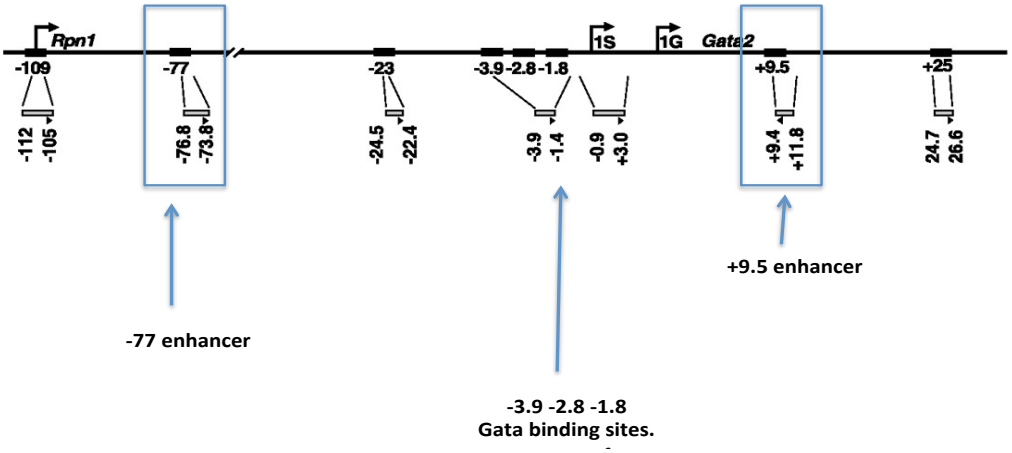


Figure 6.1: Murine *Gata2* locus showing regulatory elements Schematic of the *Gata2* locus showing cis-acting regulatory regions. 1S = haematopoietic promoter. Numbers represent distance in Kb from the transcriptional start site. Enhancer elements are indicated with blue arrows. Adapted from Grass et al. 2006 with permission.

Theoretically, *Gata2* over-expression could contribute to the MkP expansion phenotype noted in combined Gata1s/Rad21 mutant cells. In erythroid progenitors, ectopic expression of *Gata2* blocks terminal differentiation and promotes erythroblast proliferation (Briegel et al. 1993). Conversely, in a primitive myeloid cell line ectopic *Gata2* expression resulted in megakaryocytic differentiation (J. Visvader et al. 1993), while overexpression of *Gata2* in WT mouse bone marrow progenitors was associated with increased CD41+ cells and decreased polyploidisation (Huang et al. 2009). Perhaps most significantly, *Gata2* expression is increased in ML-DS patient samples compared to non DS AMKL (Bourquin et al. 2005).

It has been shown that *Gata2* is aberrantly expressed in Gata1s mutant mice (Li et al. 2005, Ling et al. 2019). Recently, Ling et al. used CUT&RUN-seq to demonstrate that Gata1s binds to the -3.9 kb, -2.8 kb, -1.8 kb, and +9.5

kb regulatory regions in erythroid cells as effectively as full-length Gata1. In the Gata1s model, however, these sites showed increased chromatin accessibility and loss of repressive H2k27 methylation, suggesting a possible mechanism of de-repression (Ling et al. 2019).

There is also evidence that depletion of cohesin may influence the *Gata2* locus. In an shRNA knockdown of *Rad21* in CD34+ cells, Mazumdar et al. reported enrichment for GATA2 binding motifs in regions of the genome showing increased accessibility compared to WT. Increased GATA2 binding to these sites was confirmed by ChIP-seq and knockdown of GATA2 in human HSPCs shown to reverse the differentiation block caused by cohesin depletion (Mazumdar et al. 2015).

In keeping with these findings, my data show that *Gata2* expression is increased in both Gata1s and combined Gata1s/Rad21 MkP, although only the Comb vs WT reached statistical significance. Genes differentially expressed between both the Combined mutant and wild type MkP and Combined mutant and Gata1s only MkP are enriched for Gata2 targets and the Gata2 transcription factor binding motif is enriched at the differentially accessible sites between combined mutant and WT control. To explore this further, I visualised the *Gata2* locus, together with downloaded and normalised Gata1/Gata1s ChIPseq data from G1ME cells (Chlon et al. 2015), Encode CTCF and Rad21 ChIP seq data from mouse bone marrow and ATAC-seq tracks from my analysis of the MkP (Figure 6.2). There is a suggestion that Gata1s (black) binding may be reduced compared to Gata1 full length (green) at the -77 enhancer, although this has not been quantitated and the impact on *Gata2* expression has not been explored. Additionally, a highly conserved forward CTCF peak (red box) is noted between the -77 enhancer and *Gata2* promoter. This appears to bind Rad21 and it is tempting to hypothesise that it might serve as a boundary element, separating the upstream enhancer from its promoter region. Depletion of Rad21 at this site could therefore influence enhancer-promoter interactions and (directly or indirectly) further impact *Gata2* gene expression. Clearly functional studies and correlation with gene expression and epigenetic data would be needed to investigate further.

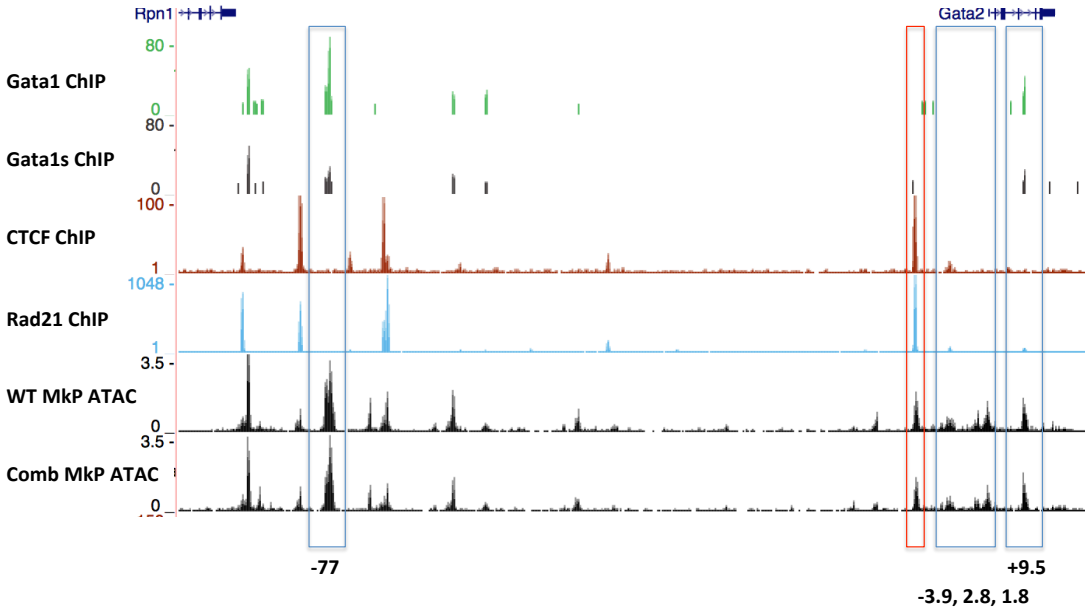


Figure 6.2: Murine *Gata2* locus showing ChIP-seq and ATAC-seq data. Visualisation of the murine Gata2 locus showing regulatory elements (blue boxes), CTCF binding site (red box) and downloaded ChIP-seq and ATAC-seq datasets as described.

6.4.2 Runx1

Another key haematopoietic regulator, *Runx1*, is widely expressed throughout the haematopoietic lineages, with the exception of mature erythroid cells where levels are low (North et al. 2004). Although essential for haematopoietic development (Okuda et al. 1996), it does not appear necessary for maintenance of HSCs in adult mice and may play a role in negative regulation of HSCs (Ichikawa et al. 2004, Ichikawa et al. 2008). Significantly, Runx1 is required for terminal megakaryopoiesis and platelet production, and *Runx1-/-* mice, are thrombocytopenic with reduced megakaryocyte ploidy (Ichikawa et al. 2004). *RUNX1* is one of the most frequently mutated genes in human AML, with the majority of reported mutations or translocations resulting in loss of function. Target genes include regulation of cell cycle and apoptosis-related genes, suggesting a mechanism for increased proliferation in cells with aberrant RUNX1 function. In humans *RUNX1* is located on Chromosome 21. Interestingly, however, in ML-DS patient samples, *RUNX1* expression was found to be decreased rather than increased compared to non-DS AMKL (Bourquin et al. 2005).

At a molecular level, Runx1 has been shown by co-immunoprecipitation to physically interact with Gata1. In one study is was proposed that these interactions depended on Gata1 N- and C-terminals but not the zinc finger domains (Elagib et al. 2003). However this was queried in a later report showing physical interaction of all investigated patient-specific GATA1 mutants with RUNX1 (Xu et al., 2006). Recently, upregulation of *Runx1* was reported in Gata1s mutant erythroid cells, associated with aberrant trimethylation of H3K27 and limiting terminal erythroid differentiation (Ling et al. 2019). This contrasts with my MkP data and the downregulation of *RUNX1* seen in ML-DS and other AMLs, and may reflect differing roles for Runx1 in erythroid and megakaryocyte progenitors.

Runx1 has also been reported to interact with cohesin. In a zebrafish model, RNAi-mediated depletion of rad21 resulted in loss of expression of runx1 and a failure in mature blood cell differentiation. This was partially rescued by microinjection of runx1 mRNA, suggesting a role for rad21 in regulation of runx1 expression (Horsfield et al. 2007). In adult AML, *RUNX1* and cohesin gene mutations were

found to significantly co-occur (Thota et al. 2014), although the mechanism of interaction remains unknown.

My data indicates that dysregulation of *Runx1* may contribute to the observed phenotype. However how Gata1s and cohesin co-operate to cause this is less clear. As reported in ML-DS, *Runx1* expression was decreased in the combined Gata1s/Rad21 mutant MkP and chromatin regions differentially accessible between combined mutant and WT were enriched for Runx1 binding motifs. Interestingly however, this enrichment was most significant in genomic regions with increased chromatin accessibility in the combined mutant - perhaps suggesting a role for additional pathways and regulatory mechanisms at this locus.

6.4.3 MTA3

As a final example, I include the metastasis associated protein-3 (MTA3) which was upregulated in the combined mutant. MTA3 is a member of the NURD (nucleosome remodelling and histone deacetylation) complex (Fujita et al., 2003). The NURD complexes function primarily as transcriptional repressors and have both chromatin remodelling and histone deacetylation properties (Basta and Rauchman, 2015 for review). The NURD complex has been shown to play critical roles in erythroid and megakaryocyte differentiation. It is recruited to Gata1-bound genes by its interaction with FOG-1 (Hong et al., 2005). It has also been shown to regulate haematopoietic stem cell self renewal (Yoshida et al., 2008). Although the exact role of MTA3 in the NURD complex is unknown, it is reported at increased levels in invasive breast cancer and B-cell lymphoma (Jaye et al., 2007).

Little has been published on interactions between cohesin and MTA3. Interestingly, however, analysis of the mouse *Mta3* locus (Figure 6.3) reveals that *Mta3* (upregulated in the Comb mutant) is located immediately adjacent to the *Haao* (3-Hydroxyanthranilate 3,4-dioxygenase) gene which is significantly downregulated in the combined mutant. Given the location of CTCF binding sites across the MTA3 locus, information regarding chromatin structure, TAD and sub-TAD localisation

would be potentially highly relevant to better understand this pattern of gene expression.

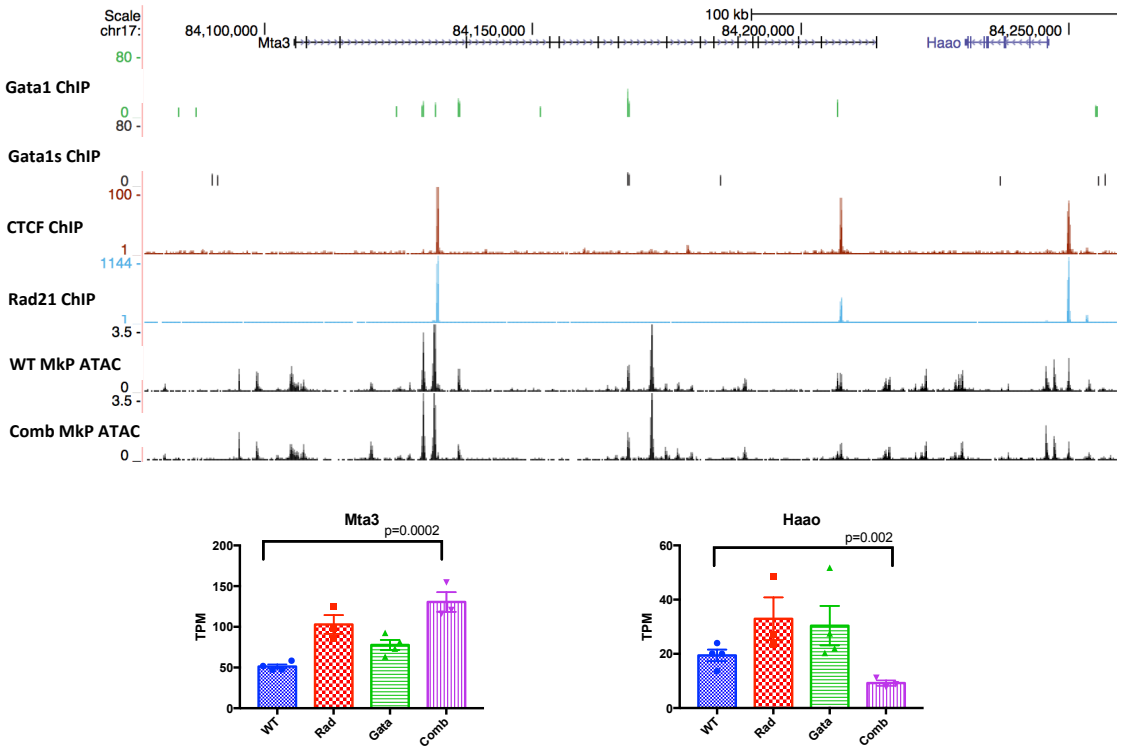


Figure 6.3: Murine *MTA3* locus showing ChIP-seq and ATAC-seq data. Visualisation of the murine MTA3 locus. The Haao gene locus is immediately downstream. Bottom plot shows gene expression for MTA3 and Haao genes in sorted MkP. Graphs show Transcript Per Million (TPM) for WT, Rad, Gata and Cohesin mutant mice. Each dot represents a biological replicate. Bar graph shows mean +/- SEM. Statistically significant differences are marked (One-way ANOVA, p-value shown)

Once again, the mechanism of co-operation seems likely to be complex - possibly involving the additive effects of differential binding of Gata1s to its co-factors and subtle changes in transcriptional structure at the MTA3 gene locus.

6.5 Conclusions

Understanding how somatic mutations co-operate to cause cancer is important but challenging, potentially involving the dysregulation and interaction of multiple cellular processes. Indeed, one emerging model of tumourigenesis is that culmination of multiple subtle perturbations in a cell or population of cells together result in significant proliferation and differentiation block.

In this thesis I have explored the interaction between N-terminal truncating mutations in the erythromegakaryocytic transcription factor Gata1 and haploinsufficient mutations in one of the core cohesin components- Rad21: a combination of mutations commonly found in Myeloid Leukaemia of Down Syndrome.

Using next generation sequencing to characterise the genetic landscape of ML-DS and its preleukaemic phase, TAM, I first confirmed the co-existence of Gata1s and cohesin gene mutations in ML-DS in a large patient cohort and characterised the frequency and predicted functional impact of these mutations on the cell. To understand how these mutations might work together to perturb haematopoiesis, I used a transgenic mouse model system, comparing Gata1s, haploinsufficiency of the core cohesin component Rad21 and combined Gata1s/Rad21 mutations. This revealed a novel haematopoietic phenotype characterised by marked expansion of the CD9+ megakaryocyte progenitor compartment and abnormal proliferation and differentiation of cells within it. Integrated molecular analysis demonstrated that Gata1s plus cohesin haploinsufficiency resulted in altered gene expression and chromatin accessibility. The magnitude of this change was greater than the effect of either mutation alone, supporting a model of co-operativity. Assessing my data in the light of published studies, it seems likely that Gata1s acts via differential binding compared to full length Gata1, resulting in either direct or indirect changes in the regulation of genes. By contrast, cohesin mutations may alter chromatin structure, facilitating new interactions between distant cis-acting elements and their promoters. A number of key transcription factors and pathways were highlighted as good targets in which to study the mechanism of interaction in more detail. However, unpicking these heterogeneous molecular events remains a considerable challenge.

Clearly a number of questions remain. Why are GATA1 mutations so common in the context of T21? How important is the T21 fetal microenvironment in this process? How do T21 and GATA1s interact to cause disease? Why do mutations in the genes encoding the cohesin complex occur at a higher frequency in ML-DS than adult AML? It is hoped that greater insight into these processes will not only translate to therapeutic advances for patients with TAM and ML-DS, but also our understanding of the molecular and cellular events underlying leukaemogenesis.

7
Appendices

Contents

7.1 Amplicon resequencing

Table 7.1: Amplicon coverage per gene for targeted resequencing of TAM and ML-DS samples. Average depth of coverage is per sample for each gene and reflects paired end reads. Adapted from Labuhn et al., 2019.

Category	Gene	No. of amplicons	Average depth of coverage
Gata1	GATA1	6	2700
Cohesin	STAG2	55	1082
	RAD21	28	1511
	SMC1A	45	1671
	SMC3	55	1462
	STAG1	57	1299
	CTCF	28	1732
	NIBPL	118	1212
Epigenetic	EZH2	32	1801
	BCOR	14	1730
	KANSL1	40	1927
	SUZ12	33	1074
	KDM6A	49	1282
	KMT2A	42	1640
	KMT2C	37	1679
	KMT2D	4	1591
	EP300	15	1545
	KAT8	1	2035
	TET2	14	1667
	DNMT3A	1	1388
	PHF8	12	1972
Tyrosine kinase	DCAF7	12	1982
	JAK1	48	1933
	JAK2	52	1210
	JAK3	49	1630
	MPL	1	582
	SH2B3	22	1357
	PDGFRB	1	1681
	FLT3	6	1766
	GNB1	6	1867
	PTEN	6	1838
	PTPN11	7	1861
	KIT	2	1878
	CBL	6	2189
	PTPRD	12	1669
RAS	NRAS	3	1144
	KRAS	4	1128
	NF1	30	1549
Transcription	ATRX	4	1424

Table 7.1 continued from previous page

Factor	RUNX1	8	1597
	ASXL1	7	1487
	CREBBP	24	1772
	CUX1	11	1778
	MYC	4	2321
Other	CSF2RB	5	2817
	SRSF2	1	736
	ADRA2C	1	0
	CSF2	2	1372
	DLEC1	5	1911
	TGD	1	667
	ZRSR2	1	1579
	NOTCH2	7	2238
	EXT1	3	1888
	MN1	1	1253
	GFPT2	1	487
	IL3	4	2449
	ITGB1	10	1114
	SETBP1	7	2432
	IDH1	2	552.3
	IDH2	2	2277
	SF3B1	5	1903
	WT1	1	2561

7.2 Antibodies used for flow cytometric sorting and analysis and viability

Table 7.2: List of antibodies used for flow cytometric sorting and analysis. *Secondary stain with Streptavidin PETxRed

Antibody	Clone	Fluorochrome	Company	Assay
B220	RA3-6B2	PECy5	BioLegend	HSC/Myeloid progenitors
c-Kit (CD117)	2B8	APC-eF780	eBiosciences	HSC
c-Kit (CD117)	2B8	APC	BioLegend	In vitro analysis of MkP
CD105	MJ7/18	Biotin	BioLegend	Myeloid progenitors
CD150	TC15-12F12.2	PECy7	BioLegend	HSC
CD150	TC15-12F12.2	APC	BioLegend	Myeloid progenitors
CD16/32	93	PE	eBiosciences	Myeloid progenitors
CD41	MWReg30	PE	BioLegend	In vitro analysis of MkP
CD41	MWReg30	PeCy7	eBiosciences	Myeloid progenitors
CD42b	Xia.G5	FITC	Emfret	In vitro analysis of MkP
CD48	HM48-1	APC	BioLegend	HSC
CD4	RM4-5	PECy5	BD	HSC/Myeloid progenitors
CD5	53-73	PECy5	BD	HSC/Myeloid progenitors
CD8a	53-6.7	PECy5	BioLegend	HSC/Myeloid progenitors
CD9	KMC8	BV650	BD	Myeloid progenitors
Flt-3	A2F10	PE	BioLegend	HSC
Gr1	RB6-8C5	PECy5	BioLegend	HSC/Myeloid progenitors
Ploidy		Hoechst 33342	Molecular Probes	In vitro analysis of MkP
Sca-1	E13-161.7	Pacific Blue	BioLegend	HSC
Ter119	TER-119	FITC	eBiosciences	Myeloid progenitors
Ter119	TER-119	PECy5	BioLegend	HSC
Ter119	TER-119	PECy7	BD	In vitro analysis of MkP
Viabiilty		7AAD	Sigma	All panels

7.3 Top 50 most highly expressed genes in WT MkP

Table 7.3: Top 50 most highly expressed genes in WT MkP samples

Name	Description
Tmsb4x	thymosin, beta 4, X chromosome
Pf4	platelet factor 4
B2m	beta-2 microglobulin
Mir682	microRNA 682
Pfn1	profilin 1
Gpx1	glutathione peroxidase 1
Rpl32	ribosomal protein L32
Rap1b	RAS related protein 1b
Atpif1	ATPase inhibitory factor 1
Ifitm2	interferon induced transmembrane protein 2
Rps11	ribosomal protein S11
Rpl4	ribosomal protein L4
Eif5a	eukaryotic translation initiation factor 5A
Cnbp	cellular nucleic acid binding protein
Rps5	ribosomal protein S5
Atp5b	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit
Rps14	ribosomal protein S14
Ifitm3	interferon induced transmembrane protein 3
Eef1a1	eukaryotic translation elongation factor 1 alpha 1
Cfl1	cofilin 1, non-muscle
Calm1	calmodulin 1
Calm3	calmodulin 3
Calm2	calmodulin 2
Slc25a5	solute carrier family 25 (mitochondrial carrier, adenine nucleotide translocator), member 5
Cox8a	cytochrome c oxidase subunit 8A
Srgn	serglycin

Cox6b1	cytochrome c oxidase, subunit 6B1
Rpl41	ribosomal protein L41
Etfb	electron transferring flavoprotein, beta polypeptide
Rpl8	ribosomal protein L8
Rps9	ribosomal protein S9
Cox4i1	cytochrome c oxidase subunit 4I1
Ndufa4	Ndufa4, mitochondrial complex associated
Ppp1ca	protein phosphatase 1 catalytic subunit alpha
Ppp1ca	protein phosphatase 1 catalytic subunit alpha
Arhgdib	Rho, GDP dissociation inhibitor (GDI) beta
Snrpb	small nuclear ribonucleoprotein B
Cd9	CD9 antigen
Cox6a1	cytochrome c oxidase subunit 6A1
Sh3bgrl3	SH3 domain binding glutamic acid-rich protein-like 3
Ldha	lactate dehydrogenase A
Fth1	ferritin heavy polypeptide 1
Actb	actin, beta
Car2	carbonic anhydrase 2
Atp5g3	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C3 (subunit 9)
Nop10	NOP10 ribonucleoprotein
Rps27l	ribosomal protein S27-like
Ranbp1	RAN binding protein 1
Ranbp1	RAN binding protein 1
Hspa8	heat shock protein 8

7.4 Differentially expressed gene lists

Table 7.4: Genes significantly differentially expressed between Comb and WT. Negative Log2fold change = down in Comb. Positive Log2fold change = up in Comb. Only genes with p-adjusted value <0.05 and Log2fold change <-0.5 and >0.5 are shown.

Name	L2FC	padj	Name	L2FC	padj
Timp3	-1.927801872	9.23E-13	Ss18	0.505959473	0.022692852
Cited4	-1.370042661	4.33E-06	Blmh	0.506545101	0.037887807
Klf1	-1.352973125	9.09E-06	Uqcc	0.507191465	0.023143571
Ddah2	-1.306342357	1.17E-05	Dgkz	0.510517653	0.043989976
Vwf	-1.273216261	1.91E-05	Rapsn	0.510891593	0.039688533
Tfec	-1.201130027	0.000126159	Tpm1	0.515826932	0.009335119
Apoe	-1.198357059	1.64E-13	Ankrd10	0.516432003	0.014924279
Sdpr	-1.194915267	1.68E-08	Rit1	0.516661404	0.036604407
Pf4	-1.114574479	3.34E-09	Ccdc40	0.516851074	0.041809227
Abcb4	-1.056153882	0.000808421	Oxr1	0.519713852	0.041037936
Nupr1	-1.049446226	0.001339844	Arhgap1	0.521561343	0.01858107
Plp2	-1.047934219	0.000122151	Nt5dc2	0.52332518	0.032691301
Cd55	-1.045693534	4.43E-06	Srd5a3	0.524096555	0.041043874
Tmem50b	-1.030284964	0.00016918	Stard7	0.524216495	0.015916288
Rgcc	-0.978647431	0.001121445	Elmo1	0.525898894	0.041822796
Sla2	-0.975455661	0.000577794	Trmt6	0.526559942	0.002776888
Trim47	-0.955981115	0.00153294	Mcm3	0.526810621	0.047469239
Mns1	-0.936071479	0.002466719	Nup62	0.528190207	0.048171859
Lpar6	-0.928797072	0.000820296	Gp5	0.528353331	0.0381484
5730508B09Rik	-0.915098178	0.002161322	Scarb1	0.528370671	0.023782832
Itih5	-0.890767858	0.004296572	Slc35b2	0.528836783	0.03111731
Bex4	-0.882951754	0.009263281	Glud1	0.529668982	0.009142674
Blvrb	-0.882683109	0.002169118	Atxn10	0.532695454	0.004037193
Mt1	-0.86180125	0.004037193	Mcm5	0.533122284	0.035242952
Nefh	-0.856548255	0.012779637	Spop	0.53329185	0.018073257
1190002N15Rik	-0.842076951	0.009335119	Noc2l	0.534372367	0.048422755
Ldhb	-0.823880181	0.017651241	Pkm	0.535427741	0.016251156
Gstm1	-0.821800253	0.000927655	Ivd	0.536582296	0.041272429
Tmsb10	-0.820463524	0.000184648	Ctsd	0.539833231	0.019872669
Vps13d	-0.819511298	0.013775164	Igfbp4	0.540706736	0.003476584
Ddx17	-0.819489966	0.005486802	Gpr108	0.550251965	0.028731084
Rpl3	-0.819004378	0.00048522	Mapk14	0.550292314	0.010185428
Pde10a	-0.806419098	0.021195927	Cxxc5	0.553273205	0.023200274
Pop5	-0.801109635	0.000199665	Timeless	0.55545646	0.03569925
Cst3	-0.796553382	0.00077332	Mta2	0.556621158	0.022995323
Hspe1	-0.796104004	2.88E-05	Parp1	0.557587529	0.016251156
Ifitm1	-0.791703264	0.000819768	Pank2	0.557872047	0.012565454
Rplp0	-0.786545417	5.93E-05	Rpa1	0.559544721	0.033735563
Grtp1	-0.779311449	0.0211235	Hemgn	0.56896677	0.009255302
Atpif1	-0.772315698	2.12E-07	Ddx20	0.569277419	0.020944807
Lgals1	-0.766604648	0.019970464	Gnpat	0.575758646	0.040038345

Table 7.4 continued from previous page

Esyt2	-0.757489775	0.021039203	Serinc3	0.576236723	9.38E-05
Csgalnact1	-0.75573087	0.006024216	Tap2	0.578485121	0.028552
Galnt6	-0.750235324	0.031414844	Jak1	0.584830487	0.032109748
D1Ertd622e	-0.748576372	0.036153045	Edc3	0.585564963	0.042745303
Tnfaip2	-0.742144066	0.037015452	Dennd1c	0.587066831	0.035501381
Ubc	-0.731397169	0.01430005	Tubb6	0.588399748	0.038518383
Smim11	-0.725921801	0.005864623	Pgam1	0.589312646	0.024628571
Rpl37	-0.725894536	0.000352475	Mcm6	0.590605559	0.01430005
Rcn3	-0.720790276	0.005486802	Map3k7	0.593795634	0.048042687
9530082P21Rik	-0.710902095	0.023200274	Cd44	0.594498063	0.017490147
Hint3	-0.710573085	0.015700928	Acadl	0.594941419	0.001788223
Gdf3	-0.703400151	0.047874023	Coro1c	0.595791742	0.038518383
Ifitm3	-0.701887887	0.000310923	Bin1	0.595821094	0.020860119
Ylpm1	-0.697074271	0.03038118	Gpr56	0.599884027	0.019056638
Cox7c	-0.690660294	0.004699708	Wars	0.60205958	0.00261902
Mdm1	-0.688931104	0.000801974	Txnrd1	0.602149468	0.025674466
Usmg5	-0.685628758	0.000983711	Tmem161a	0.602925079	0.040621092
Pbx1	-0.683884534	0.003736949	Parn	0.603858558	0.032109748
Dynll1	-0.677733555	0.022256923	Snx4	0.604035567	0.030327261
Rnasek	-0.675964505	0.000297759	Uba1	0.604504163	0.043488121
Id2	-0.673027443	0.048100313	Sirt3	0.60551424	0.004440053
Tomm6	-0.669613933	0.000126159	Btk	0.607622927	0.028069927
Hmgcs1	-0.669162574	0.049417527	Pgm2	0.608281775	0.027407647
Whsc1l1	-0.668404904	0.028485876	Stxbp2	0.612298608	0.012833119
Gm13826	-0.665332516	0.000310923	1600014C10Rik	0.614051412	0.048100313
1110008P14Rik	-0.66206083	0.047874023	Anapc2	0.614376343	0.009335119
Atp5k	-0.661285231	0.002162392	Flot1	0.619325243	0.001373673
Krtcap2	-0.660146595	2.75E-05	Anapc5	0.62035289	0.002379534
Macrod2	-0.655042753	0.023782832	Rtel1	0.621833541	0.041037936
Rangrf	-0.652887879	0.01619004	Ubash3a	0.622121385	0.047874023
Rplp2	-0.652460874	0.001373271	9030624J02Rik	0.622516701	0.025677219
Ms4a3	-0.646919135	0.023249226	Lrmp	0.623258883	0.000353753
Lamtor3	-0.64656073	0.000529199	Acly	0.624719807	0.042046349
Sepw1	-0.646108629	0.001357346	Poldip2	0.625071693	5.79E-06
Mrpl52	-0.643942651	0.013946587	Cpsf1	0.626164409	0.03378579
Ndufaf4	-0.638451174	0.00272211	Nt5dc1	0.627258829	0.047469239
Arsk	-0.636333787	0.003515223	Slain1	0.628528186	0.041043874
Timm8b	-0.636154254	0.008198871	Rmnd5b	0.629800626	0.021213818
Polr1d	-0.635352366	0.000141342	Arhgap15	0.631363715	0.008587628
Cox17	-0.632989885	0.000571746	Glt25d1	0.633020885	0.006862444
Rps21	-0.632013798	0.005023278	Fam174b	0.633791206	0.020328212
Fam195a	-0.629930268	0.012873343	Egln2	0.637222428	0.045081017
Uqcrq	-0.629772687	5.93E-05	Vwa5a	0.638933349	0.008366708
Reep2	-0.627430472	0.039329325	Zfp384	0.638989455	0.014677924
Mrps6	-0.625243284	0.003580269	Klhl6	0.639653412	0.001450412
Rps26	-0.624340714	0.002770647	Slc38a10	0.640313953	0.043488121
Crip1	-0.622975705	0.001834583	Vps33b	0.644998907	0.036615135

Table 7.4 continued from previous page

Gm6402	-0.621852056	0.00749757	Gata2	0.647618154	0.016159833
Icam2	-0.619654026	0.006436369	Vars	0.647775024	0.040310777
Pbx4	-0.619652367	0.048100313	Lepre1	0.651858783	0.040038345
Slirp	-0.618949742	0.005097482	Ikbkap	0.653971514	0.036173591
Egfl7	-0.618377202	0.027412344	Grk6	0.654969422	0.00770285
Pdzk1ip1	-0.617989937	0.000228665	Gga1	0.659144581	0.023995195
Rps3a1	-0.617486862	0.032733624	D930015E06Rik	0.661538768	0.033946189
Snrpg	-0.615589277	0.002429711	Lonp2	0.665416888	0.01011247
Dnajc15	-0.612952591	0.000744006	Axin1	0.665657353	0.029820379
Rpl35	-0.604447915	0.00261902	Git2	0.67115851	0.000649184
Mrpl41	-0.604409758	0.000297759	Pigq	0.672529418	0.008982209
Zfml	-0.603537012	0.01889978	Sh2d5	0.673114036	0.045105993
Fkbp11	-0.600725505	0.018466866	Wdr45	0.673949819	0.039144008
Ndufa13	-0.599840921	3.21E-05	Amdhd2	0.674063607	0.043410985
Cisd1	-0.599823744	1.17E-05	Smad4	0.675362584	0.041243893
Anapc13	-0.599693058	0.001757331	Frat2	0.675625226	0.040310777
Capza1	-0.598423385	0.043488121	Gle1	0.675930562	0.032521493
Ndufa5	-0.59775408	0.003453767	Pdxdc1	0.677295612	0.025801512
Nenf	-0.594314084	0.014677924	Gchfr	0.679318931	0.048100313
Cox20	-0.59354492	0.012425957	Inpp5b	0.680884264	1.07E-05
Ccdc107	-0.593326958	0.014244922	Itga2b	0.681755722	0.021487752
Fam178a	-0.592939413	0.037665383	Sla	0.684869302	0.000297759
Rpl36al	-0.592850621	0.000218726	Hbp1	0.688117656	0.020408089
Gm11974	-0.592659451	0.000126159	Ankrd39	0.689123612	0.011437887
Shfm1	-0.591357189	0.00011801	Cybasc3	0.690094832	0.04829346
Cox6a1	-0.590069346	0.000118302	Pafah2	0.691797506	0.001686318
Atp5j2	-0.588599531	0.00235832	Rbpms2	0.695772893	0.010735239
Rpl39	-0.57688373	0.010633903	Mfn2	0.709200434	0.041796347
Atp5e	-0.576616163	0.00272211	Cd180	0.711279331	0.037094409
Cox6b2	-0.574411092	0.014420481	Klc4	0.711798001	0.0381484
Rps28	-0.573449015	0.010860717	Ptpn7	0.712006221	0.007255935
Tmem256	-0.573320952	0.001245044	Stil	0.712082496	0.037665383
Ubl5	-0.572159631	0.002835734	Rab3gap2	0.71301376	0.03705517
Med29	-0.572079262	0.018771099	Vps45	0.713321528	0.007365727
Tomm7	-0.571627812	0.000349077	Rnf26	0.719430912	0.038416052
Lsm5	-0.57160981	0.01266963	Rwdd2b	0.722399022	0.028619496
Gm15706	-0.570594896	0.048578352	Arhgap9	0.722691771	0.00781246
Acyp1	-0.568146365	0.009335119	Ubxn11	0.72820789	0.038568528
2010003O02Rik	-0.567871557	0.012687833	Nckap1l	0.733592681	0.010661062
Nudt8	-0.566780056	0.021039203	Crbn	0.740422438	0.004037193
Bola3	-0.565726827	0.003223664	Tusc1	0.741453881	0.032566478
Ptprcap	-0.562912254	1.65E-05	Ppp1r16a	0.743563713	0.03705517
Mrpl20	-0.560184337	0.000727422	Cmtm6	0.743644141	0.019975598
Vamp5	-0.558445489	0.002190448	Cd27	0.744319844	5.97E-05
Cox6c	-0.556065381	0.00182066	Add1	0.744722498	0.006078798
2010107E04Rik	-0.555198797	0.001173039	Lrrc41	0.745055074	0.000311489
Hsp90aa1	-0.554631681	0.014995377	Gosr2	0.747136662	0.000134954

Table 7.4 continued from previous page

Cacybp	-0.553767741	0.023130788	Cd1d1	0.748504155	0.012687833
Vimp	-0.55239827	0.040377724	Poldip3	0.749489522	0.01889978
Rpl13	-0.551325138	0.000927655	Ankrd13a	0.752131048	0.000171559
Ndufb3	-0.551098195	0.000796682	St7	0.753681042	0.013652594
Rps15	-0.54926231	0.000145476	Rwdd3	0.755103833	0.019435141
1190007I07Rik	-0.546437687	0.011437887	Pcmtd2	0.760249637	0.000225045
Dpm1	-0.543455851	0.028619496	Sh2b2	0.761315826	0.030977218
BC002163	-0.543273092	0.011668959	Atp2a3	0.761889076	0.009557909
Pet100	-0.542764305	0.009255302	Zfp346	0.764367859	0.030694043
Tceb2	-0.542629003	0.000927655	Prkcq	0.765607127	3.65E-06
Tmem258	-0.54253061	0.000139162	Lgals8	0.770006943	0.000571746
Cox7a2	-0.542308767	0.002629179	Tubb2a	0.770432998	0.029820379
Chchd1	-0.541651092	0.007109069	Pfkl	0.772916996	0.006059724
2010012O05Rik	-0.540830353	0.001120258	Il16	0.774358685	0.004226377
Romo1	-0.540227423	0.003204257	Ccdc122	0.777801876	0.014244922
Coa6	-0.539817211	0.012309851	Pecam1	0.779023983	0.015238443
Rpl35a	-0.538350197	0.03223127	Hace1	0.78002628	0.021195927
Rps27l	-0.537433299	0.001077498	A930005H10Rik	0.780897869	0.016272958
Rpl32	-0.537417588	0.001357346	Cd34	0.781581879	0.010670222
Gm5148	-0.53637125	0.000126159	Mfge8	0.784441408	0.000819768
Snrpb2	-0.536187856	0.008366708	Cst7	0.788422041	0.005023278
Btf3	-0.535634058	0.010801509	Arhgap18	0.789000929	0.000543489
Rpl37a	-0.535043634	0.020101847	Klhl12	0.789204937	0.01430005
Myeov2	-0.534180124	0.002145977	Nsmf	0.789757877	0.021025559
Bola2	-0.533793423	0.002770647	Mfsd2b	0.800686855	2.32E-06
Ndufa1	-0.533322549	0.009335119	Prune	0.812348657	0.01961462
Rpl23	-0.533253353	0.027407647	Dhrs3	0.823317142	0.003852365
Ndufa6	-0.532430948	0.003563148	Il20ra	0.826056604	0.017092718
Mrpl33	-0.532190956	0.002629179	Mtmr14	0.826155017	9.06E-06
Ube2f	-0.529872982	0.012838779	Vps52	0.826618516	0.00272211
Rps10	-0.528877613	0.007630178	Casp2	0.827714556	0.00235832
2310057M21Rik	-0.528639053	0.045081017	Zfp46	0.833323975	0.012873343
Mrpl30	-0.52761896	0.014924279	Fam189b	0.839163646	0.009557909
Mnf1	-0.527381747	0.007307807	Acp2	0.843076698	0.006600976
Cd9	-0.525616383	0.000821177	Car3	0.848262317	0.011083694
Rwdd1	-0.524540112	0.001031166	Lrrc40	0.848760004	0.001645995
Timm9	-0.522452956	0.03012063	Bex6	0.848951256	0.0126522
Rpl7a	-0.52163828	0.034493036	Taf1a	0.850891974	0.013652594
Rpl21	-0.521433322	0.014677924	Trim35	0.852576585	0.000111834
Ndufc1	-0.521101091	0.028121016	Cyhr1	0.863244614	0.003453767
Manf	-0.521097706	0.047469239	Dip2b	0.864356101	0.000773263
Rps24	-0.520637326	0.02649534	Dcaf8	0.869297641	0.001098103
Rplp1	-0.519886224	0.005486802	Abtb1	0.869556227	0.000353753
Cdk2ap2	-0.519830166	0.000185126	Usp28	0.875427064	0.010801509
Ndufc2	-0.51933671	0.001233602	Pls3	0.880974198	0.005486802
Hmgn3	-0.519306813	0.020081432	Bckdha	0.88118953	0.002379534
Rexo2	-0.519255987	0.002886556	Smpd4	0.886113655	0.002769702

Table 7.4 continued from previous page

Magoh	-0.516655889	0.00488723	Fcer1g	0.891584478	0.001283345
Osbpl8	-0.51646801	0.046359021	Tspan14	0.894419198	2.06E-06
Ndufab1	-0.515648844	0.02649534	Rnf180	0.906524486	0.001233602
Hypk	-0.512475533	0.002718902	Lat2	0.92248331	0.002737872
Atp5o	-0.512166515	0.001357346	Pomgnt1	0.928637445	0.003764787
Ndufb6	-0.511613148	0.000735168	Ccdc22	0.93026939	0.000184648
Ccdc124	-0.509872431	0.014621277	1810026J23Rik	0.941820851	0.001515105
Gcsh	-0.509739186	0.00048522	Ints7	0.944547812	0.000577794
Minos1	-0.50922778	0.004948647	Slc28a2	0.949340106	0.004564002
Rpl19	-0.507662887	0.020007257	Zyx	0.958263994	4.43E-06
Rpl6	-0.506880493	0.006708949	Ung	0.976627781	0.002655565
Rps8	-0.506521679	0.009335119	Pstpip2	0.983233094	0.003086872
Cox6b1	-0.506417619	0.001957273	Rnpep	0.988816521	5.93E-05
Znhit1	-0.505463022	0.009084467	Pcbd1	1.00315106	0.00048522
Ndufs6	-0.505443015	0.004081923	Cd96	1.021084448	0.000634392
Snhg5	-0.50539125	0.030738949	Arhgap6	1.031925085	0.000425953
Rps7	-0.504899339	0.009335119	Dis3l2	1.037849419	0.000225045
Ndufa3	-0.50473942	0.003764787	Rab3d	1.048974579	3.27E-05
Mir682	-0.503927967	0.005486802	Emid1	1.056570773	0.000139162
Chchd7	-0.503504683	0.011668959	Bcas3	1.063965541	0.000571746
Pfdn1	-0.501969633	0.025599171	Gpr97	1.064216508	7.51E-26
Fam96b	-0.501861569	0.020081432	Mta3	1.069742756	6.39E-08
Higd1a	-0.501507396	0.012873343	P2ry14	1.085383239	0.000819768
Edf1	-0.500672725	0.001930074	Gpr171	1.127410338	7.02E-05
Rpl13a	-0.500331498	0.001139321	Cd300a	1.147554592	2.14E-05
Eif3g	-0.500297319	0.005900488	Tceanc	1.16207855	9.38E-05
			Dfna5	1.201761925	5.00E-05
			Clec4d	1.231945298	1.62E-05
			Ncf4	1.244877012	8.69E-05
			Engase	1.287238178	1.17E-05

Table 7.5: Genes significantly differentially expressed between Gata and WT. Negative Log2fold change = down in Gata. Positive Log2fold change = up in Gata. Only genes with p-adjusted value <0.05 and Log2fold change <-0.5 and >0.5 are shown

Name	L2FC	padj	Name	L2FC	padj
Klf1	-1.415407808	3.54E-06	Nol11	0.501430689	0.04461827
Gm15915	-1.386235662	1.85E-05	Hp1bp3	0.509391234	0.0004567
Timp3	-1.243721806	5.74E-05	Poldip2	0.51536652	0.000226267
5730508B09Rik	-1.128125732	0.000120641	Plin3	0.515747383	0.042687856
Id2	-1.083547058	0.001811558	Sar1a	0.516683829	0.039137215
Cd24a	-1.074350808	0.002795669	Wars	0.522648317	0.02221962
Blvrb	-0.947724272	0.001576724	Pde4b	0.524854744	0.041188333
Pf4	-0.931521065	1.13E-06	Fbxl6	0.52873851	0.039137215
Lpar6	-0.884256338	0.002834434	Ctsd	0.539437186	0.034353438
Lgals1	-0.883531529	0.022759356	Stard7	0.543835914	0.02221962
Bex4	-0.876758957	0.03180032	Parp1	0.545168649	0.034166615
Cited4	-0.856460461	0.02221962	Mipep	0.574066738	0.036092963
Clu	-0.850502611	0.033873273	Sirt3	0.584429225	0.011974257
Hmgcs1	-0.840356158	0.022759356	Acadl	0.600398371	0.002242689
Aff3	-0.83758657	0.034108934	Ivd	0.602871482	0.034108934
Itgb7	-0.83015562	0.034166615	Slc43a3	0.605661912	0.023288004
Endod1	-0.806644637	0.02221962	Ddx54	0.606308807	0.015343208
Tfrc	-0.805264876	0.027568783	Dgkz	0.609264936	0.023015758
Sla2	-0.76944513	0.023015758	Usb1	0.611070942	0.027197025
A630089N07Rik	-0.769228263	0.033873273	Mcm6	0.611508007	0.02221962
Tmem98	-0.760106396	0.033873273	Rapsn	0.612160804	0.021819077
Rreb1	-0.734049574	0.046497518	Vwa5a	0.619037598	0.02221962
Aplp2	-0.733017187	0.026383689	Wdr5	0.629524553	0.004160675
Hook3	-0.726453526	0.015343208	Dennd1c	0.631808736	0.040984075
Plp2	-0.719071402	0.031753576	Mtmr14	0.63352844	0.001541938
Rps3a1	-0.662163169	0.038439257	Mfsd2b	0.646326144	0.000155632
Zfml	-0.658952527	0.01885729	Add1	0.649051427	0.039736247
0610030E20Rik	-0.658285064	0.02221962	Casp2	0.650576808	0.048327974
Apoe	-0.648508894	0.000502501	Tsc22d3	0.654287424	0.00218113
Snhg3	-0.646185183	0.027931445	Cd44	0.657111865	0.014407844
Tmsb10	-0.64611263	0.009701477	Abtb1	0.662813597	0.02221962
Dnajc24	-0.644737035	0.023288004	Prkcq	0.664554088	5.73E-05
Cst3	-0.598750009	0.033873273	Map3k7	0.665184168	0.045351992
Cox7c	-0.598650838	0.034108934	Rnpep	0.67612307	0.022759356
S100a11	-0.587579788	0.03965621	Surf6	0.678929144	0.026383689
Pbx1	-0.585079847	0.03232044	Serpine2	0.68747675	0.031190936
Rasgrp2	-0.583402251	0.000152528	Dyrk3	0.715765162	0.005852133
Mif4gd	-0.579914836	0.02326217	Il10rb	0.716852751	0.00247894
Ndufaf4	-0.566902034	0.020354941	Mta2	0.71763092	0.002555447
Cln3	-0.562142448	0.02221962	Trim35	0.721593151	0.002225914
Degs1	-0.543693915	5.73E-05	Inpp5b	0.728154761	7.78E-07
Ptprcap	-0.514033937	6.87E-05	Gata2	0.740294251	0.008537385
Fra10ac1	-0.510319733	0.030332159	Stx11	0.749994874	0.022759356
			BC052040	0.750477509	0.04846763

Table 7.5 continued from previous page

Ankrd13a	0.760539616	0.000120641
Mcoln1	0.765479988	0.003827662
Ece1	0.766827274	0.028611106
Pold1	0.77094542	0.042687856
Cd300a	0.778267391	0.01885729
Rnf180	0.778815761	0.01885729
Fam174b	0.783392445	0.00412568
Ccdc22	0.802263132	0.002795669
Bap1	0.808503377	0.041821847
Engase	0.814019699	0.029137024
Emid1	0.815922974	0.011974257
Cln8	0.832043031	0.018383
Rab3d	0.833884335	0.002555447
Slc39a11	0.839194741	0.033873273
Tprn	0.844291457	0.02221962
Tap2	0.845696431	0.000522468
Gpr97	0.86442907	5.33E-19
E130308A19Rik	0.87286838	0.030332159
Gorasp1	0.895457504	0.023015758
Pigq	0.962752063	5.70E-05
Clec4d	0.973293497	0.002242689
A930005H10Rik	1.005172408	0.002266257
Stat2	1.006746607	0.007964062
Rgs1	1.015076963	0.000629232
Bcas3	1.015422302	0.00247894
Il20ra	1.017735248	0.00689501
Pomgnt1	1.027719209	0.002471845
Dis3l2	1.084102441	0.000137846
Dfna5	1.129035811	0.000226267
Ints7	1.22680063	2.39E-06
Tnfaip3	1.240506853	0.000242906

Table 7.6: Genes significantly differentially expressed between Rad and WT. NegativeLog2fold change = down in Rad. Positive Log2fold change = up in Rad. Only genes with p-adjusted value <0.05 and Log2fold change <-0.5 and >0.5 are shown.

Name	**L2FC**	**padj**	**Name**	**L2FC**	**padj**
Cited4	-1.950379101	1.09E-12	0610031J06Rik	0.531301551	0.013356392
Mt1	-1.218093314	7.50E-05	Flot1	0.539982433	0.034796869
Pf4	-0.884641368	7.11E-05	Acadl	0.573501024	0.016289063
Hsph1	-0.863621899	0.032962268	Inpp5b	0.616781311	0.000436418
Abcb4	-0.856280922	0.048270144	Igfbp4	0.654582478	0.001294904
Ralgps2	-0.841668899	0.031418558	Timeless	0.690089785	0.030160311
Mdm1	-0.807669138	0.000330231	Pde4b	0.693680461	0.009284421
B3galt6	-0.802764406	0.027094451	Abtb1	0.696540608	0.032366245
Sdpr	-0.683442093	0.023305443	Ctla2a	0.710809252	0.03816277
Cd55	-0.683109948	0.028769874	Mcoln1	0.715828598	0.027782226
Ndufaf4	-0.641343812	0.016289063	Pafah2	0.728694018	0.005611502
Cd9	-0.616388541	0.000330231	Sdr39u1	0.754507652	0.036658812
Pdcd4	-0.606360503	0.022274813	Pnrc1	0.763382367	0.016289063
Hspe1	-0.601279373	0.016289063	Marcksl1	0.765949652	0.031324334
Scand1	-0.556092978	0.027782226	Rtel1	0.797121499	0.028334739
Bola2	-0.534458657	0.016289063	Mta3	0.821064276	0.000436418
			Tmem184c	0.873764056	0.027782226
			Ints7	0.888540565	0.009625976
			Cd69	0.920530346	0.031418558
			P2ry14	0.928011884	0.030160311
			Ccdc122	0.929864492	0.015279851
			Fos	0.955354667	0.019149827
			Dfna5	1.004522253	0.006438298
			Pcbd1	1.020561484	0.002229948
			Tnfaip3	1.086834959	0.005611502
			Cd34	1.087905699	0.00067688
			Rab3d	1.100779937	7.11E-05

References

Adolfsson, J., R. Månsson, N. Buza-Vidas, A. Hultquist, K. Liuba, C. T. Jensen, D. Bryder, L. Yang, O.-J. Borge, L. A. Thoren, et al. (2005). "Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential". *Cell* 121.2, pp. 295–306.

Ahmed, M., A. Sternberg, G. Hall, A. Thomas, O. Smith, A. O'Marcaigh, R. Wynn, R. Stevens, M. Addison, D. King, et al. (2004). "Natural history of GATA1 mutations in Down syndrome". *Blood* 103.7, pp. 2480–2489.

Akashi, K., D. Traver, T. Miyamoto, and I. L. Weissman (2000). "A clonogenic common myeloid progenitor that gives rise to all myeloid lineages". *Nature* 404.6774, pp. 193–197.

Alford, K. A., K. Reinhardt, C. Garnett, A. Norton, K. Böhmer, C. Von Neuhoff, A. Kolenova, E. Marchi, J. H. Klusmann, I. Roberts, et al. (2011). "Analysis of GATA1 mutations in down syndrome transient myeloproliferative disorder and myeloid leukemia". *Blood* 118.8, pp. 2222–2238.

Alford, K. A., A. Slender, L. Vanes, Z. Li, E. M. C. Fisher, D. Nizetic, S. H. Orkin, I. Roberts, and V. L. J. Tybulewicz (2010). "Perturbed hematopoiesis in the Tc1 mouse model of Down syndrome." *Blood* 115.14, pp. 2928–37.

Banno, K., S. Omori, K. Hirata, K. Wada, Y. Kitabatake, K. O. Correspondence, N. Nawa, N. Nakagawa, K. Nishimura, M. Ohtaka, et al. (2016). "Systematic Cellular Disease Models Reveal Synergistic Interaction of Trisomy 21 and GATA1 Mutations in Hematopoietic Abnormalities". *Cell Reports* 15.

Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, et al. (2012). "NCBI GEO: archive for functional genomics data sets—update". *Nucleic Acids Research* 41.D1, pp. D991–D995.

Bartley, T., J. Bogenberger, P. Hunt, Y.-S. Li, H. Lu, F. Martin, M.-S. Chang, B. Samal, J. Nichol, S. Swift, et al. (1994). "Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor MpI". *Cell* 77.7, pp. 1117–1124.

Birger, Y., L. Goldberg, T. M. Chlon, B. Goldenson, I. Muler, G. Schiby, J. Jacob-Hirsch, G. Rechavi, J. D. Crispino, and S. Izraeli (2013). "Perturbation of fetal hematopoiesis in a mouse model of Down syndrome's transient myeloproliferative disorder." *Blood* 122.6, pp. 988–98.

Blobel, G. A., C. A. Sieff, and S. H. Orkin (1995). "Ligand-dependent repression of the erythroid transcription factor GATA-1 by the estrogen receptor." *Molecular and cellular biology* 15.6, pp. 3147–53.

Boer, J. de, A. Williams, G. Skavdis, N. Harker, M. Coles, M. Tolaini, T. Norton, K. Williams, K. Roderick, A. J. Potocnik, et al. (2003). *Transgenic mice with hematopoietic and lymphoid specific expression of Cre.*

Bolger, A. M., M. Lohse, and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". *Bioinformatics* 30.15, pp. 2114–2120.

Bourquin, J.-P., A. Subramanian, C. Langebrake, D. Reinhardt, O. Bernard, P. Ballerini, A. Baruchel, H. Lè Ne Cavé, N. Dastugue, H. Hasle, et al. (2005). "Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling".

Bresnick, E. H., K. R. Katsumura, H.-Y. Lee, K. D. Johnson, and A. S. Perkins (2012). "Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies". *Nucleic Acids Research* 40.13, pp. 5819–5831.

Briegel, K., K. C. Lim, C. Plank, H. Beug, J. D. Engel, and M. Zenke (1993). "Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner." *Genes & development* 7.6, pp. 1097–109.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." *Nature methods* 10.12, pp. 1213–8.

Buenrostro, J. D., B. Wu, H. Y. Chang, and W. J. Greenleaf (2015). "ATAC-seq: A method for assaying chromatin accessibility genome-wide". *Current Protocols in Molecular Biology* 2015, pp. 1–21.

Byrska-Bishop, M., D. VanDorn, A. E. Campbell, M. Betensky, P. R. Arca, Y. Yao, P. Gadue, F. F. Costa, R. L. Nemiroff, G. A. Blobel, et al. (2015). "Pluripotent stem cells reveal erythroid-specific activities of the GATA1 N-terminus". *The Journal of Clinical Investigation* 125.3, pp. 993–1005.

Calligaris, R., S. Bottardi, S. Cogoi, I. Apezteguia, and C. Santoro (1995). "Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor." *Proceedings of the National Academy of Sciences of the United States of America* 92.25, pp. 11598–602.

Cantor, A. B. and S. H. Orkin (2002). "Transcriptional regulation of erythropoiesis: an affair involving multiple partners". *Oncogene* 21.21, pp. 3368–3376.

Carrelha, J., y. Meng, L. M. Kettyle, T. C. Luis, R. Norfo, V. Alcolea, H. boukarabila, f. Grasso, A. Gambardella, A. Grover, et al. (2018). "Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells". *Nature Publishing Group*.

Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. Meirelles, N. R. Clark, and A. Ma'ayan (2013). "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool". *BMC Bioinformatics* 14.1, p. 128.

Chlon, T. M., M. McNulty, B. Goldenson, A. Rosinski, and J. D. Crispino (2015). "Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression." *Haematologica* 100.5, pp. 575–84.

Chou, S. T., E. Khandros, L. C. Bailey, K. E. Nichols, C. R. Vakoc, Y. Yao, Z. Huang, J. D. Crispino, R. C. Hardison, G. A. Blobel, et al. (2009). "Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate." *Blood* 114.5, pp. 983–94.

Consortium, T. E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489.7414, pp. 57–74.

Corces, M. R. (2015). "full-text". *Nature Genetics*.

Corces, M. R., J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, et al. (2016). "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution". *Nature Genetics*.

Craddock, C., S. Tauro, P. Moss, and D. Grimwade (2005). "Biology and management of relapsed acute myeloid leukaemia". *British Journal of Haematology* 129.1, pp. 18–34.

Crispino, J. D. (2005). "GATA1 in normal and malignant hematopoiesis". *Seminars in Cell & Developmental Biology* 16.1, pp. 137–147.

Crispino, J. D., M. B. Lodish, J. P. MacKay, and S. H. Orkin (1999). "Use of Altered Specificity Mutants to Probe a Specific Protein–Protein Interaction in Differentiation: the GATA-1:FOG Complex". *Molecular Cell* 3.2, pp. 219–228.

Cuartero, S., F. D. Weiss, G. Dharmalingam, Y. Guo, E. Ing-Simmons, S. Masella, I. Robles-Rebollo, X. Xiao, Y.-F. Wang, I. Barozzi, et al. (2018). "Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation". *Nature Immunology* 19.9, pp. 932–941.

Davis, C. A., B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, et al. (2018). "The Encyclopedia of DNA elements (ENCODE): data portal update." *Nucleic acids research* 46.D1, pp. D794–D801.

Denker, A. and W. d. Laat (2016). "The second decade of 3C technologies: detailed insights into nuclear organization". *Genes & Development* 30.12, p. 1357.

Despang, A., R. Schöpflin, M. Franke, S. Ali, I. Jerković, C. Paliou, W.-L. Chan, B. Timmermann, L. Wittler, M. Vingron, et al. (2019). "Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture". *Nature Genetics* 51.8, pp. 1263–1271.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions". *Nature* 485.7398, pp. 376–380.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics (Oxford, England)* 29.1, pp. 15–21.

Doré, L. C., J. D. Crispino, N. Debili, L. Coulombel, L. Croisille, A. Vannucchi, F. Paoletti, S. Linari, K. Akashi, D. Traver, et al. (2011). "Transcription factor networks in erythroid cell and megakaryocyte development." *Blood* 118.2, pp. 231–9.

Drissen, R., N. Buza-Vidas, P. Woll, S. Thongjuea, A. Gambardella, A. Giustacchini, E. Mancini, A. Zriwil, M. Lutteropp, A. Grover, et al. (2016). "Distinct myeloid progenitor differentiation pathways identified through single cell RNA sequencing". *Nature immunology* 17.6, p. 666.

Elagib, K. E., F. K. Racke, M. Mogass, R. Khetawat, L. L. Delehanty, and A. N. Goldfarb (2003). "RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation". *Blood* 101.11, pp. 4333–4341.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller (2016). "MultiQC: summarize analysis results for multiple tools and samples in a single report". *Bioinformatics* 32.19, pp. 3047–3048.

Fazio, G., V. Massa, A. Grioni, V. Bystry, S. Rigamonti, C. Saitta, M. Galbiati, C. Rizzari, C. Consarino, A. Biondi, et al. (2019). "First evidence of a paediatric patient with Cornelia de Lange syndrome with acute lymphoblastic leukaemia." *Journal of clinical pathology* 72.8, pp. 558–561.

Fox, A. H., C. Liew, M. Holmes, K. Kowalski, J. Mackay, and M. Crossley (1999). "Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers". *The EMBO Journal* 18.10, pp. 2812–2822.

Frontelo, P., D. Manwani, M. Galdass, H. Karsunky, F. Lohmann, P. G. Gallagher, and J. J. Bieker (2007). "Novel role for EKLF in megakaryocyte lineage commitment". *Blood* 110.12, pp. 3871–3880.

Fujiwara, Y., C. P. Browne, K. Cunniff, S. C. Goff, and S. H. Orkin (1996). "Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1." *Proceedings of the National Academy of Sciences of the United States of America* 93.22, pp. 12355–8.

Gamis, A. S., T. A. Alonzo, R. B. Gerbing, J. M. Hilden, A. D. Sorrell, M. Sharma, T. W. Loew, R. J. Arceci, D. Barnard, J. Doyle, et al. (2011). "Natural history of transient myeloproliferative disorder clinically diagnosed in Down syndrome neonates: a report from the Children's Oncology Group Study A2971." *Blood* 118.26, pp. 6752–9.

Gan, T., C. D. Jude, K. Zaffuto, and P. Ernst (2010). "Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis." *Leukemia* 24.10, pp. 1732–41.

Gandhi, R., P. J. Gillespie, and T. Hirano (2006). "Human Wapl Is a Cohesin-Binding Protein that Promotes Sister-Chromatid Resolution in Mitotic Prophase". *Current Biology* 16.24, pp. 2406–2417.

Gao, X., K. D. Johnson, Y.-I. Chang, M. E. Boyer, C. N. Dewey, J. Zhang, and E. H. Bresnick (2013). "Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo." *The Journal of experimental medicine* 210.13, pp. 2833–42.

Goldfarb, A. N. (2009). "Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb." *Journal of cellular biochemistry* 107.3, pp. 377–82.

Grass, J. A., M. E. Boyer, S. Pal, J. Wu, M. J. Weiss, and E. H. Bresnick (2003). "GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling." *Proceedings of the National Academy of Sciences of the United States of America* 100.15, pp. 8811–6.

Grass, J. A., H. Jing, S.-I. Kim, M. L. Martowicz, S. Pal, G. A. Blobel, and E. H. Bresnick (2006). "Distinct functions of dispersed GATA factor complexes at an endogenous gene locus." *Molecular and cellular biology* 26.19, pp. 7056–67.

Groet, J., S. McElwaine, M. Spinelli, A. Rinaldi, I. Burtscher, C. Mulligan, A. Mensah, S. Cavani, F. Dagna-Bricarelli, G. Basso, et al. (2003). "Acquired mutations in GATA1 in neonates with Down's syndrome with transient myeloid disorder." *Lancet (London, England)* 361.9369, pp. 1617–20.

Grove, C. S. and G. S. Vassiliou (2014). "Acute myeloid leukaemia: a paradigm for the clonal evolution of cancer?" *Disease models & mechanisms* 7.8, pp. 941–51.

Gruber, T. A., A. Larson Gedman, J. Zhang, C. S. Koss, S. Marada, H. Q. Ta, S.-C. Chen, X. Su, S. K. Ogden, J. Dang, et al. (2012). "An Inv(16)(p13.3q24.3)-Encoded CBFA2T3-GLIS2 Fusion Protein Defines an Aggressive Subtype of Pediatric Acute Megakaryoblastic Leukemia". *Cancer Cell* 22.5, pp. 683–697.

Gutiérrez, L., T. Nikolic, T. B. van Dijk, H. Hammad, N. Vos, M. Willart, F. Grosveld, S. Philipsen, and B. N. Lambrecht (2007). "Gata1 regulates dendritic-cell development and survival." *Blood* 110.6, pp. 1933–41.

Halsey, C., O. Tunstall, B. Gibson, I. Roberts, and G. Graham (2010). "Role of GATA-1s in early hematopoiesis and differences between alternative splicing in human and murine GATA-1". *Blood* 115.16, pp. 3415–3416.

Hasle, H., I. H. Clemmensen, and M. Mikkelsen (2000). "Risks of leukaemia and solid tumours in individuals with Down's syndrome." *Lancet (London, England)* 355.9199, pp. 165–9.

Heidinger-Pauli, J. M., O. Mert, C. Davenport, V. Guacci, and D. Koshland (2010). "Systematic Reduction of Cohesin Differentially Affects Chromosome Segregation, Condensation, and DNA Repair". *Current Biology* 20.10, pp. 957–963.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass (2010). "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities". *Molecular Cell* 38.4, pp. 576–589.

Herault, Y., J. M. Delabar, E. M. C. Fisher, V. L. J. Tybulewicz, E. Yu, and V. Brault (2017). "Rodent models in Down syndrome research: impact and future opportunities." *Disease models & mechanisms* 10.10, pp. 1165–1186.

Hill, V. K., J.-S. Kim, and T. Waldman (2016). "Cohesin mutations in human cancer." *Biochimica et biophysica acta* 1866.1, pp. 1–11.

Hitzler, J. K., J. Cheung, Y. Li, S. W. Scherer, and A. Zipursky (2003). "GATA1 mutations in transient leukemia and acute megakaryoblastic leukemia of Down syndrome." *Blood* 101.11, pp. 4301–4.

Hollanda, L. M., C. S. P. Lima, A. F. Cunha, D. M. Albuquerque, J. Vassallo, M. C. Ozelo, P. P. Joazeiro, S. T. O. Saad, and F. F. Costa (2006). "An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis". *Nature Genetics* 38.7, pp. 807–812.

Hong, W., M. Nakazawa, Y.-Y. Chen, R. Kori, C. R. Vakoc, C. Rakowski, and G. A. Blobel (2005). "FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1". *The EMBO Journal* 24.13, pp. 2367–2378.

Horsfield, J. A., S. H. Anagnostou, J. K.-H. Hu, K. H. Y. Cho, R. Geisler, G. Lieschke, K. E. Crosier, and P. S. Crosier (2007). "Cohesin-dependent regulation of Runx genes". *Development* 134.14, pp. 2639–2649.

Huang, Z., L. C. Dore, Z. Li, S. H. Orkin, G. Feng, S. Lin, and J. D. Crispino (2009). "GATA-2 reinforces megakaryocyte development in the absence of GATA-1." *Molecular and cellular biology* 29.18, pp. 5168–80.

Ichikawa, M., T. Asai, T. Saito, G. Yamamoto, S. Seo, I. Yamazaki, T. Yamagata, K. Mitani, S. Chiba, H. Hirai, et al. (2004). "AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis". *Nature Medicine* 10.3, pp. 299–304.

Ichikawa, M., S. Goyama, T. Asai, M. Kawazu, M. Nakagawa, M. Takeshita, S. Chiba, S. Ogawa, and M. Kurokawa (2008). "AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis." *Journal of immunology (Baltimore, Md. : 1950)* 180.7, pp. 4402–8.

Jacobsen, S. E. W. and C. Nerlov (2019). "Haematopoiesis in the era of advanced single-cell technologies". *Nature Cell Biology* 21.1, pp. 2–8.

Johnson, K. D., G. Kong, X. Gao, Y.-I. Chang, K. J. Hewitt, R. Sanalkumar, R. Prathibha, E. A. Ranheim, C. N. Dewey, J. Zhang, et al. (2015). "Cis-regulatory mechanisms governing stem and progenitor cell transitions." *Science advances* 1.8, e1500503.

Kadri, Z., R. Shimizu, O. Ohneda, L. Maouche-Chretien, S. Gisselbrecht, M. Yamamoto, P.-H. Romeo, P. Leboulch, and S. Chretien (2009). "Direct Binding of pRb/E2F-2 to GATA-1 Regulates Maturation and Terminal Cell Division during Erythropoiesis". *PLoS Biology* 7.6. Ed. by M. A. Goodell, e1000123.

Kagey, M. H., J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture". *Nature* 467.7314, pp. 430–435.

Kaushansky, K., S. Lok, R. D. Holly, V. C. Broudy, N. Lin, M. C. Bailey, J. W. Forstrom, M. M. Buddle, P. J. Oort, F. S. Hagen, et al. (1994). "Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin". *Nature* 369.6481, pp. 568–571.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler (2002). "The human genome browser at UCSC." *Genome research* 12.6, pp. 996–1006.

Kiel, M. J., Ö. H. Yilmaz, T. Iwashita, O. H. Yilmaz, C. Terhorst, and S. J. Morrison (2005). "SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells". *Cell* 121.7, pp. 1109–1121.

Kirsammer, G., S. Jilani, H. Liu, E. Davis, S. Gurbuxani, M. M. Le Beau, and J. D. Crispino (2008). "Highly penetrant myeloproliferative disease in the Ts65Dn mouse model of Down syndrome." *Blood* 111.2, pp. 767–75.

Klusmann, J.-H., U. Creutzig, M. Zimmermann, M. Dworzak, N. Jorch, C. Langebrake, A. Pekrun, K. Macakova-Reinhardt, and D. Reinhardt (2008). "Treatment and prognostic impact of transient leukemia in neonates with Down syndrome." *Blood* 111.6, pp. 2991–8.

Klusmann, J.-H., F. J. Godinho, K. Heitmann, A. Maroz, M. L. Koch, D. Reinhardt, S. H. Orkin, and Z. Li (2010). "Developmental stage-specific interplay of GATA1 and IGF signaling in fetal megakaryopoiesis and leukemogenesis." *Genes & development* 24.15, pp. 1659–72.

Koboldt, D. C., D. E. Larson, and R. K. Wilson (2013). "Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection." *Current protocols in bioinformatics* 44, pp. 1–17.

Kon, A., L.-Y. Shih, M. Minamino, M. Sanada, Y. Shiraishi, Y. Nagata, K. Yoshida, Y. Okuno, M. Bando, R. Nakato, et al. (2013). "Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms". *Nature Genetics* 45.10, pp. 1232–1237.

Kondo, M., I. L. Weissman, and K. Akashi (1997). "Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow". *Cell* 91.5, pp. 661–672.

Korbel, J. O., T. Tirosh-Wagner, A. E. Urban, X.-N. Chen, M. Kasowski, L. Dai, F. Grubert, C. Erdman, M. C. Gao, K. Lange, et al. (2009). "The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies." *Proceedings of the National Academy of Sciences of the United States of America* 106.29, pp. 12031–6.

Koulnis, M., R. Pop, E. Porpiglia, J. R. Shearstone, D. Hidalgo, and M. Socolovsky (2011). "Identification and analysis of mouse erythroid progenitors using the CD71/TER119 flow-cytometric assay." *Journal of visualized experiments : JoVE* 54.

Kuhl, C., A. Atzberger, F. Iborra, B. Nieswandt, C. Porcher, and P. Vyas (2005). "GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1." *Molecular and cellular biology* 25.19, pp. 8592–606.

Kuvardina, O. N., J. Herglotz, S. Kolodziej, N. Kohrs, S. Herkt, B. Wojcik, T. Oellerich, J. Corso, K. Behrens, A. Kumar, et al. (2015). "RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation." *Blood* 125.23, pp. 3570–9.

Labuhn, M., K. Perkins, S. Matzk, L. Varghese, C. Garnett, E. Papaemmanuil, M. Metzner, A. Kennedy, V. Amstislavskiy, T. Risch, et al. (2019). "Mechanisms of Progression of Myeloid Preleukemia to Transformed Myeloid Leukemia in Children with Down Syndrome". *Cancer Cell*.

Lai, A. Y. and M. Kondo (2006). "Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors." *The Journal of experimental medicine* 203.8, pp. 1867–73.

Lambert, M. P., L. G. Jackson, D. Clark, M. Kaur, I. D. Krantz, and M. A. Deardorff (2011). "The incidence of thrombocytopenia in children with Cornelia de Lange syndrome". *American Journal of Medical Genetics, Part A* 155.1, pp. 33–37.

Langebrake, C., U. Creutzig, and D. Reinhardt (2005). "Immunophenotype of Down Syndrome Acute Myeloid Leukemia and Transient Myeloproliferative Disease Differs Significantly from Other Diseases with Morphologically Identical or Similar Blasts". *Klinische Pädiatrie* 217.3, pp. 126–134.

Langmead, B. (2010). "Aligning short sequencing reads with Bowtie". *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* CHAPTER, Unit.

Laurenti, E. and B. Göttgens (2018). "From haematopoietic stem cells to complex differentiation landscapes." *Nature* 553.7689, pp. 418–426.

Leinonen, R., H. Sugawara, M. Shumway, and International Nucleotide Sequence Database Collaboration (2011). "The sequence read archive." *Nucleic acids research* 39.Database issue, pp. 19–21.

Leı, A., L. Perié, K. R. Duffy, L. Kok, R. J. de Boer, and T. N. Schumacher Correspondence (2015). "The Branching Point in Erythro-Myeloid Differentiation". *Cell* 163.

Letourneau, A., F. A. Santoni, X. Bonilla, M. R. Sailani, D. Gonzalez, J. Kind, C. Chevalier, R. Thurman, R. S. Sandstrom, Y. Hibaoui, et al. (2014). "Domains of genome-wide gene expression dysregulation in Down's syndrome". *Nature* 508.7496, pp. 345–350.

Ley, T. J. (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." *The New England journal of medicine* 368.22, pp. 2059–74.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1. G. P. D. P. 1000 Genome Project Data Processing Subgroup (2009). "The Sequence Alignment/Map format and SAMtools." *Bioinformatics (Oxford, England)* 25.16, pp. 2078–9.

Li, Z., F. J. Godinho, J.-H. Klusmann, M. Garriga-Canut, C. Yu, and S. H. Orkin (2005). "Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1". *Nature genetics* 37.6, pp. 613–619.

Liao, Y., G. K. Smyth, and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". *Bioinformatics* 30.7, pp. 923–930.

Ling, T., Y. Birger, M. J. Stankiewicz, N. Ben-Haim, T. Kalisky, A. Rein, E. Kugler, W. Chen, C. Fu, K. Zhang, et al. (2019). "Chromatin occupancy and epigenetic analysis reveal new insights into the function of GATA1 N-terminus in erythropoiesis." *Blood*, blood.2019001234.

Liu, J. and I. D. Krantz (2009). "Cornelia de Lange syndrome, cohesin, and beyond". *Clin Genet* 76.4, pp. 303–314.

Lordier, L., D. Bluteau, A. Jalil, C. Legrand, J. Pan, P. Rameau, D. Jouni, O. Bluteau, T. Mercher, C. Leon, et al. (2012). "RUNX1-induced silencing of non-muscle myosin heavy chain IIB contributes to megakaryocyte polyploidization". *Nature Communications* 3.1, p. 717.

Losada, A. (2014). "Cohesin in cancer: chromosome segregation and beyond". *Nature Reviews Cancer* 14.6, pp. 389–393.

Losada, A., T. Yokochi, and T. Hirano (2005). "Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts." *Journal of cell science* 118.Pt 10, pp. 2133–41.

Loughran, S. J., E. A. Kruse, D. F. Hacking, C. A. de Graaf, C. D. Hyland, T. A. Willson, K. J. Henley, S. Ellis, A. K. Voss, D. Metcalf, et al. (2008). "The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells". *Nature Immunology* 9.7, pp. 810–819.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biology* 15.12, p. 550.

Luc, S., K. Anderson, S. Kharazi, N. Buza-Vidas, C. Böiers, C. T. Jensen, Z. Ma, L. Wittmann, and S. E. W. Jacobsen (2008). "Down-regulation of Mpl marks the transition to lymphoid-primed multipotent progenitors with gradual loss of granulocyte-monocyte potential." *Blood* 111.7, pp. 3424–34.

Lunter, G. and M. Goodson (2011). "Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads." *Genome research* 21.6, pp. 936–9.

Malinge, S., M. Bliss-Moreau, G. Kirsammer, L. Diebold, T. Chlon, S. Gurbuxani, and J. D. Crispino (2012). "Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome." *The Journal of clinical investigation* 122.3, pp. 948–62.

Malinge, S., T. Chlon, L. C. Doré, R. P. Ketterling, M. S. Tallman, E. Paietta, A. S. Gamis, J. W. Taub, S. T. Chou, M. J. Weiss, et al. (2013). "Development of acute megakaryoblastic leukemia in Down syndrome is associated with sequential epigenetic changes." *Blood* 122.14, pp. 33–43.

Mancini, E., A. Sanjuan-Pla, L. Luciani, S. Moore, A. Grover, A. Zay, K. D. Rasmussen, S. Luc, D. Bilbao, D. O'Carroll, et al. (2012). "FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors." *The EMBO journal* 31.2, pp. 351–65.

Martin, D. I. and S. H. Orkin (1990). "Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1". *Genes and Development*.

Martowicz, M. L., J. A. Grass, M. E. Boyer, H. Guend, and E. H. Bresnick (2005). "Dynamic GATA factor interplay at a multicomponent regulatory region of the GATA-2 locus." *The Journal of biological chemistry* 280.3, pp. 1724–32.

Massey, G. V., A. Zipursky, M. N. Chang, J. J. Doyle, S. Nasim, J. W. Taub, Y. Ravindranath, G. Dahl, and H. J. Weinstein (2006). "A prospective study of the natural history of transient leukemia (TL) in neonates with Down syndrome (DS): Children's Oncology Group (COG) study POG-9481". *Blood*.

Mazumdar, C., Y. Shen, S. Xavy, F. Zhao, A. Reinisch, R. Li, M. R. Corces, R. A. Flynn, J. D. Buenrostro, S. M. Chan, et al. (2015). "Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation". *Cell Stem Cell* 17.6, pp. 675–688.

McDevitt, M. A., R. A. Shivdasani, Y. Fujiwara, H. Yang, and S. H. Orkin (1997). "A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1." *Proceedings of the National Academy of Sciences of the United States of America* 94.13, pp. 6781–5.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, et al. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." *Genome research* 20.9, pp. 1297–303.

Merkenschlager, M. and E. P. Nora (2016). "CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation". *Annual Review of Genomics and Human Genetics* 17.1, pp. 17–43.

Merkenschlager, M. and D. T. Odom (2013). "CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets". *Cell* 152.6, pp. 1285–1297.

Miccio, A., Y. Wang, W. Hong, G. D. Gregory, H. Wang, X. Yu, J. K. Choi, S. Shelat, W. Tong, M. Poncz, et al. (2010). "NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development". *The EMBO Journal* 29.2, pp. 442–456.

Migliaccio, A. R., R. A. Rana, M. Sanchez, R. Lorenzini, L. Centurione, L. Bianchi, A. M. Vannucchi, G. Migliaccio, and S. H. Orkin (2003). "GATA-1 as a Regulator of Mast Cell Differentiation Revealed by the Phenotype of the GATA-1low Mouse Mutant". *Journal of Experimental Medicine* 197.3, pp. 281–296.

Molkentin, J. D. (2000). "The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression." *The Journal of biological chemistry* 275.50, pp. 38949–52.

Morrison, S., A. Wandycz, H. Hemmati, D. Wright, and I. Weissman (1997). "Identification of a lineage of multipotent hematopoietic progenitors". *Development* 124.10.

Mouthon, M. A., O. Bernard, M. T. Mitjavila, P. H. Romeo, W. Vainchenker, and D. Mathieu-Mahul (1993). "Expression of tal-1 and GATA-binding proteins during human hematopoiesis". *Blood* 81.3, pp. 647–655.

Mullenders, J., B. Aranda-Orgilles, P. Lhoumaud, M. Keller, J. Pae, K. Wang, C. Kayembe, P. P. Rocha, R. Raviram, Y. Gong, et al. (2015). "Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms". *Journal of Experimental Medicine* 212.11, pp. 1833–1850.

Muntean, A. G. and J. D. Crispino (2005). "Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development".

Muramatsu, H., K. Kato, N. Watanabe, K. Matsumoto, T. Nakamura, Y. Horikoshi, J. Mimaya, C. Suzuki, M. Hayakawa, and S. Kojima (2008). "Risk factors for early death in neonates with Down syndrome and transient leukaemia". *British Journal of Haematology* 142.4, pp. 610–615.

Nakorn, T. N., T. Miyamoto, and I. L. Weissman (2003). "Characterization of mouse clonogenic megakaryocyte progenitors." *Proceedings of the National Academy of Sciences of the United States of America* 100.1, pp. 205–210.

Nardelli, J., D. Thiesson, Y. Fujiwara, F.-Y. Tsai, and S. H. Orkin (1999). "Expression and Genetic Interaction of Transcription Factors GATA-2 and GATA-3 during Development of the Mouse Central Nervous System". *Developmental Biology* 210.2, pp. 305–321.

Nasmyth, K. and C. H. Haering (2009). "Cohesin: Its Roles and Mechanisms". *Annual Review of Genetics* 43.October, pp. 525–558.

Nerlov, C., E. Querfurth, H. Kulessa, and T. Graf (2000). *GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription*. Tech. rep.

Ng, A. P., C. D. Hyland, D. Metcalf, C. L. Carmichael, S. J. Loughran, L. Di Rago, B. T. Kile, and W. S. Alexander (2010). "Trisomy of Erg is required for myeloproliferation in a mouse model of Down syndrome." *Blood* 115.19, pp. 3966–9.

Ng, A. P., M. Kauppi, D. Metcalf, L. Di Rago, C. D. Hyland, and W. S. Alexander (2012). "Characterization of thrombopoietin (TPO)-responsive progenitor cells in adult mouse bone marrow with in vivo megakaryocyte and erythroid potential." *Proceedings of the National Academy of Sciences of the United States of America* 109.7, pp. 2364–9.

Nikolaev, S. I., F. Santoni, A. Vannier, E. Falconnet, E. Giarin, G. Basso, A. Hoischen, J. A. Veltman, J. Groet, D. Nizetic, et al. (2013). "Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down syndrome." *Blood* 122.4, pp. 554–561.

Noetzli, L. J., S. L. French, and K. R. Machlus (2019). "New Insights Into the Differentiation of Megakaryocytes From Hematopoietic Progenitors". *Arteriosclerosis, Thrombosis, and Vascular Biology* 39.7, pp. 1288–1300.

Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, et al. (2012). "Spatial partitioning of the regulatory landscape of the X-inactivation centre." *Nature* 485.7398, pp. 381–5.

North, T. E., T. Stacy, C. J. Matheny, N. A. Speck, and M. F. d. Bruijn (2004). "Runx1 Is Expressed in Adult Mouse Hematopoietic Stem Cells and Differentiating Myeloid and Lymphoid Cells, But Not in Maturing Erythroid Cells". *Stem Cells* 22.2, pp. 158–168.

Ogawa, M., Y. Matsuzaki, S. Nishikawa, S. Hayashi, T. Kunisada, T. Sudo, T. Kina, H. Nakauchi, and S. Nishikawa (1991). "Expression and function of c-kit in hemopoietic progenitor cells." *The Journal of experimental medicine* 174.1, pp. 63–71.

Okuda, T., J. van Deursen, S. W. Hiebert, G. Grosveld, and J. R. Downing (1996). "AML1, the Target of Multiple Chromosomal Translocations in Human Leukemia, Is Essential for Normal Fetal Liver Hematopoiesis". *Cell* 84.2, pp. 321–330.

Orkin, S. H. and L. I. Zon (2008). "Hematopoiesis: An Evolving Paradigm for Stem Cell Biology". *Cell* 132.4, pp. 631–644.

Palii, C. G., Q. Cheng, M. A. Gillespie, P. Shannon, M. Mazurczyk, G. Napolitani, N. D. Price, J. A. Ranish, E. Morrissey, D. R. Higgs, et al. (2019). "Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate". *Cell Stem Cell* 24.5, pp. 812–820.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, et al. (2016). "Genomic Classification and Prognosis in Acute Myeloid Leukemia". *New England Journal of Medicine* 374.23, pp. 2209–2221.

Papaemmanuil, E., M. Gerstung, L. Malcovati, S. Tauro, G. Gundem, P. Van Loo, C. J. Yoon, P. Ellis, D. C. Wedge, A. Pellagatti, et al. (2013). "Clinical and biological implications of driver mutations in myelodysplastic syndromes." *Blood* 122.22, pp. 3616–27.

Pauli, A., F. Althoff, R. A. Oliveira, S. Heidmann, O. Schuldiner, C. F. Lehner, B. J. Dickson, and K. Nasmyth (2008). "Cell-Type-Specific TEV Protease Cleavage Reveals Cohesin Functions in Drosophila Neurons". *Developmental Cell* 14.2, pp. 239–251.

Pevny, L., M. C. Simon, E. Robertson, W. H. Klein, S.-F. Tsai, V. D'Agati, S. H. Orkin, and F. Costantini (1991). "Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1". *Nature* 349.6306, pp. 257–260.

Pietras, E. M., D. Reynaud, Y.-A. Kang, D. Carlin, F. J. Calero-Nieto, A. D. Leavitt, J. M. Stuart, B. Göttgens, and E. Passegué (2015). "Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions." *Cell stem cell* 17.1, pp. 35–46.

Pine, S. R., Q. Guo, C. Yin, S. Jayabose, C. M. Druschel, and C. Sandoval (2007). "Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome." *Blood* 110.6, pp. 2128–31.

Pronk, C. J., D. J. Rossi, R. Månsson, J. L. Attema, G. L. Norddahl, C. K. F. Chan, M. Sigvardsson, I. L. Weissman, and D. Bryder (2007). "Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy". *Cell Stem Cell* 1.4, pp. 428–442.

Psaila, B., N. Barkas, D. Iskander, A. Roy, S. Anderson, N. Ashley, V. S. Caputo, J. Lichtenberg, S. Loaiza, D. M. Bodine, et al. (2016). "Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways". *Genome Biology* 17.1, p. 83.

Quek, L., P. Ferguson, M. Metzner, I. Ahmed, A. Kennedy, C. Garnett, S. Jeffries, C. Walter, K. Piechocki, A. Timbs, et al. (2016a). "Mutational analysis of disease relapse in patients allografted for acute myeloid leukemia". *Blood Advances* 1.3, pp. 193–204.

Quek, L., G. W. Otto, C. Garnett, L. Lhermitte, D. Karamitros, B. Stoilova, I.-J. Lau, J. Doondeea, B. Usukhbayar, A. Kennedy, et al. (2016b). "Genetically distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage." *The Journal of experimental medicine* 213.8, pp. 1513–35.

Rainis, L., D. Bercovich, S. Strehl, A. Teigler-Schlegel, B. Stark, J. Trka, N. Amariglio, A. Biondi, I. Muler, G. Rechavi, et al. (2003). "Mutations in exon 2 of GATA1 are early events in megakaryocytic malignancies associated with trisomy 21." *Blood* 102.3, pp. 981–6.

Ramírez, F., F. Dündar, S. Diehl, B. A. Grüning, and T. Manke (2014). "deepTools: a flexible platform for exploring deep-sequencing data". *Nucleic Acids Research* 42.Web Server issue, W187.

Rao, A., R. K. Hills, C. Stiller, B. E. Gibson, S. S. Graaf, I. M. Hann, A. O'Marcaigh, K. Wheatley, and D. K. Webb (2006). "Treatment for myeloid leukaemia of Down syndrome: population-based experience in the UK and results from the Medical Research Council AML 10 and AML 12 trials". *British Journal of Haematology* 132.5, pp. 576–583.

Reinhardt, K., A. Kate, K. Bohmer, P. Vyas, and R. Dirk (2010). "GATA1 Mutation In Transient Leukemia (TL) and Myeloid Leukemia of Down Syndrome". *Blood* 116.21.

Remeseiro, S., A. Cuadrado, S. Kawauchi, A. L. Calof, A. D. Lander, and A. Losada (2013). "Reduction of Nipbl impairs cohesin loading locally and affects transcription but not cohesion-dependent functions in a mouse model of Cornelia de Lange Syndrome". *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1832.12, pp. 2097–2102.

Risso, D., J. Ngai, T. P. Speed, and S. Dudoit (2014). "Normalization of RNA-seq data using factor analysis of control genes or samples". *Nature Biotechnology* 32.9, pp. 896–902.

Roberts, I., K. Alford, G. Hall, G. Juban, H. Richmond, A. Norton, G. Vallance, K. Perkins, E. Marchi, S. McGowan, et al. (2013). "GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia." *Blood* 122.24, pp. 3908–17.

Rodriguez, P., E. Bonte, J. Krijgsveld, K. E. Kolodziej, B. Guyot, A. J. R. Heck, P. Vyas, E. de Boer, F. Grosveld, and J. Strouboulis (2005). "GATA-1 forms distinct activating and repressive complexes in erythroid cells". *The EMBO Journal* 24.13, pp. 2354–2366.

Roy, A., G. Cowan, A. J. Mead, S. Filippi, G. Bohn, A. Chaidos, O. Tunstall, J. K. Chan, M. Choolani, P. Bennett, et al. (2012). "Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21". *Proceedings of the National Academy of Sciences of the United States of America* 109.43, pp. 17579–17584.

Sanjuan-Pla, A., I. C. Macaulay, C. T. Jensen, P. S. Woll, T. C. Luis, A. Mead, S. Moore, C. Carella, S. Matsuoka, T. B. Jones, et al. (2013). "Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy". *Nature* 502.7470, pp. 232–236.

Sankaran, V. G., R. Ghazvinian, R. Do, P. Thiru, J.-A. Vergilio, A. H. Beggs, C. A. Sieff, S. H. Orkin, D. G. Nathan, E. S. Lander, et al. (2012). "Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia." *The Journal of clinical investigation* 122.7, pp. 2439–43.

Sasca, D., H. Yun, G. Giotopoulos, J. Szybinski, T. Evan, N. K. Wilson, M. Gerstung, P. Gallipoli, A. R. Green, R. K. Hills, et al. (2019). "Cohesin-dependent regulation of gene expression during differentiation is lost in Cohesin-mutated myeloid malignancies". *Blood*.

Sauvage, F. J. de, P. E. Hass, S. D. Spencer, B. E. Malloy, A. L. Gurney, S. A. Spencer, W. C. Darbonne, W. J. Henzel, S. C. Wong, W.-J. Kuang, et al. (1994). "Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand". *Nature* 369.6481, pp. 533–538.

Schrier, S. A., I. Sherer, M. A. Deardorff, D. Clark, L. Audette, L. Gillis, A. D. Kline, L. Ernst, K. Loomes, I. D. Krantz, et al. (2011). "Causes of death and autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature". *American Journal of Medical Genetics, Part A* 155.12, pp. 3007–3024.

Seitan, V. C., A. J. Faure, Y. Zhan, R. P. McCord, B. R. Lajoie, E. Ing-Simmons, B. Lenhard, L. Giorgetti, E. Heard, A. G. Fisher, et al. (2013). "Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments". *Genome Research* 23.12, pp. 2066–2077.

Seitan, V. C., B. Hao, K. Tachibana-Konwalski, T. Lavagnolli, H. Mira-Bontenbal, K. E. Brown, G. Teng, T. Carroll, A. Terry, K. Horan, et al. (2011). "A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation." *Nature* 476.7361, pp. 467–71.

Sherry, S. T., M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin (2001). "dbSNP: the NCBI database of genetic variation". *Nucleic Acids Research* 29.1, pp. 308–311.

Shimizu, R., S. Takahashi, K. Ohneda, J. D. Engel, and M. Yamamoto (2001). "In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis." *The EMBO journal* 20.18, pp. 5250–60.

Shivdasani, R. A., Y. Fujiwara, M. A. McDevitt, and S. H. Orkin (1997). "A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development." *The EMBO journal* 16.13, pp. 3965–73.

Shivdasani, R. A., M. F. Rosenblatt, D. Zucker-Franklin, C. W. Jackson, P. Hunt, C. J. Saris, and S. H. Orkin (1995). "Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoeitin/MGDF in megakaryocyte development". *Cell* 81.5, pp. 695–704.

Smith, L. G., I. L. Weissman, and S. Heimfeld (1991). "Clonal analysis of hematopoietic stem-cell differentiation in vivo." *Proceedings of the National Academy of Sciences of the United States of America* 88.7, pp. 2788–92.

Sofueva, S., E. Yaffe, W.-C. Chan, D. Georgopoulou, M. Vietri Rudan, H. Mira-Bontenbal, S. M. Pollard, G. P. Schroth, A. Tanay, and S. Hadjur (2013). "Cohesin-mediated interactions organize chromosomal domain architecture." *The EMBO journal* 32.24, pp. 3119–29.

Sohma, Y., H. Akahori, N. Seki, T.-a. Hori, K. Ogami, T. Kato, Y. Shimada, K. Kawamura, and H. Miyazaki (1994). "Molecular cloning and chromosomal localization of the human thrombopoietin gene". *FEBS Letters* 353.1, pp. 57–61.

Starck, J., M. Weiss-Gayet, C. Gonnet, B. Guyot, J.-M. Vicat, and F. Morlé (2010). "Inducible Fli-1 gene deletion in adult mice modifies several myeloid lineage commitment decisions and accelerates proliferation arrest and terminal erythrocytic differentiation." *Blood* 116.23, pp. 4795–805.

Stumpf, M., C. Waskow, M. Krötschel, D. van Essen, P. Rodriguez, X. Zhang, B. Guyot, R. G. Roeder, and T. Borggrefe (2006). "The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220." *Proceedings of the National Academy of Sciences of the United States of America* 103.49, pp. 18504–9.

Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, et al. (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." *Proceedings of the National Academy of Sciences of the United States of America* 101.16, pp. 6062–7.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proceedings of the National Academy of Sciences of the United States of America* 102.43, pp. 15545–50.

Takemoto, C. M., S. Brandal, A. G. Jegga, Y.-N. Lee, A. Shahlaee, Y. Ying, R. Dekoter, and M. A. McDevitt (2010). "PU.1 positively regulates GATA-1 expression in mast cells." *Journal of immunology (Baltimore, Md. : 1950)* 184.8, pp. 4349–61.

Tamblyn, J. A., A. Norton, L. Spurgeon, V. Donovan, A. Bedford Russell, J. Bonnici, K. Perkins, P. Vyas, I. Roberts, and M. D. Kilby (2016). "Prenatal therapy in transient abnormal myelopoiesis: a systematic review." *Archives of disease in childhood. Fetal and neonatal edition* 101.1, pp. 67–71.

Tate, J. G., S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, et al. (2019). "COSMIC: the Catalogue Of Somatic Mutations In Cancer". *Nucleic Acids Research* 47.D1, pp. D941–D947.

Telenius, J., T. W. Consortium, and J. R. Hughes (2018). "NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control". *bioRxiv* 16, p. 393413.

Thiollier, C., C. K. Lopez, B. Gerby, C. Ignacimouttou, S. Poglio, Y. Duffourd, J. Guégan, P. Rivera-Munoz, O. Bluteau, V. Mabialah, et al. (2012). "Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models." *The Journal of experimental medicine* 209.11, pp. 2017–31.

Thol, F., R. Bollin, M. Gehlhaar, C. Walter, M. Dugas, K. J. Suchanek, A. Kirchner, L. Huang, A. Chaturvedi, M. Wichmann, et al. (2014). "Mutations in the cohesin complex in acute myeloid leukemia: Clinical and prognostic implications". *Blood* 123.6, pp. 914–920.

Thota, S., A. D. Viny, H. Makishima, B. Spitzer, T. Radivoyevitch, B. Przychodzen, M. A. Sekeres, R. L. Levine, and J. P. Maciejewski (2014). "Genetic alterations of the cohesin complex genes in myeloid malignancies." *Blood* 124.11, pp. 1790–1798.

Toki, T., R. Kanezaki, E. Kobayashi, H. Kaneko, M. Suzuki, R. Wang, K. Terui, H. Kanegane, M. Maeda, M. Endo, et al. (2013). "Naturally occurring oncogenic GATA1 mutants with internal deletions in transient abnormal myelopoiesis in Down syndrome." *Blood* 121.16, pp. 3181–4.

Tsai, S.-F., D. I. K. Martin, L. I. Zon, A. D. D'Andrea, G. G. Wong, and S. H. Orkin (1989). "Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells". *Nature* 339.6224, pp. 446–451.

Tunstall, O., N. Bhatnagar, B. James, A. Norton, A. S. O'Marcaigh, T. Watts, A. Greenough, P. Vyas, I. Roberts, and M. Wright (2018). "Guidelines for the investigation and management of Transient Leukaemia of Down Syndrome". *British Journal of Haematology* 182.2, pp. 200–211.

Vial, Y., J. Lachenaud, A. Verloes, M. Besnard, O. Fenneteau, E. Lainey, A. Marceau-Renaut, C. Preudhomme, A. Baruchel, H. Cavé, et al. (2018). "Down syndrome-like acute megakaryoblastic leukemia in a patient with Cornelia de Lange syndrome." *Haematologica* 103.6, e274–e276.

Viny, A. D., R. L. Bowman, Y. Liu, V.-P. Lavallée, S. E. Eisman, W. Xiao, B. H. Durham, A. Navitski, J. Park, S. Braunstein, et al. (2019). "Stag1 and Stag2 regulate cell fate decisions in hematopoiesis through non-redundant topological control". *bioRxiv*, p. 581868.

Viny, A. D., C. J. Ott, B. Spitzer, M. Rivas, C. Meydan, E. Papalexi, D. Yelin, K. Shank, J. Reyes, A. Chiu, et al. (2015). "Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis". *Journal of Experimental Medicine* 212.11, pp. 1819–1832.

Visvader, J. E., M. Crossley, J. Hill, S. H. Orkin, and J. M. Adams (1995). "The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line." *Molecular and cellular biology* 15.2, pp. 634–41.

Visvader, J. and J. Adams (1993). "Megakaryocytic differentiation induced in 416B myeloid cells by GATA-2 and GATA-3 transgenes or 5-azacytidine is tightly coupled to GATA-1 expression". *Blood* 82.5.

Vyas, P., K. Ault, C. W. Jackson, S. H. Orkin, and R. A. Shivdasani (1999). "Consequences of GATA-1 Deficiency in Megakaryocytes and Platelets". *Blood* 93.9.

Wall, L., E. deBoer, and F. Grosveld (1988). "The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein." *Genes & development* 2.9, pp. 1089–100.

Walters, D. K., T. Mercher, T.-L. Gu, T. O'Hare, J. W. Tyner, M. Loriaux, V. L. Goss, K. A. Lee, C. A. Eide, M. J. Wong, et al. (2006). "Activating alleles of JAK3 in acute megakaryoblastic leukemia". *Cancer Cell* 10.1, pp. 65–75.

Wang, K., M. Li, and H. Hakonarson (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." *Nucleic acids research* 38.16, e164.

Wang, T., B. Glover, G. Hadwiger, C. A. Miller, O. di Martino, and J. S. Welch (2019). "Smc3 is required for mouse embryonic and adult hematopoiesis". *Experimental Hematology* 70, pp. 70–84.

Wang, X., J. D. Crispino, D. L. Letting, M. Nakazawa, M. Poncz, and G. A. Blobel (2002). "Control of megakaryocyte-specific gene expression by GATA-1 and FOG-1: role of Ets transcription factors". *The EMBO Journal* 21.19, pp. 5225–5234.

Wechsler, J., M. Greene, M. A. McDevitt, J. Anastasi, J. E. Karp, M. M. Le Beau, and J. D. Crispino (2002). "Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome". *Nature Genetics* 32.1, pp. 148–152.

Weiss, M. J., C. Yu, and S. H. Orkin (1997). "Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line." *Molecular and cellular biology* 17.3, pp. 1642–51.

Weiss, M. J. and S. H. Orkin (1995). "Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis". *Proceedings of the National Academy of Sciences of the United States of America* 92.21, pp. 9623–9627.

Welch, J. J., J. A. Watts, C. R. Vakoc, Y. Yao, H. Wang, R. C. Hardison, G. A. Blobel, L. A. Chodosh, and M. J. Weiss (2004). "Global regulation of erythroid gene expression by transcription factor GATA-1". *Blood* 104.10, pp. 3136–3147.

Wendt, K. S., K. Yoshida, T. Itoh, M. Bando, B. Koch, E. Schirghuber, S. Tsutsumi, G. Nagae, K. Ishihara, T. Mishiro, et al. (2008). "Cohesin mediates transcriptional insulation by CCCTC-binding factor." *Nature* 451.7180, pp. 796–801.

White, J. K., A.-k. Gerdin, N. a. Karp, E. Ryder, M. Buljan, J. N. Bussell, J. Salisbury, S. Clare, N. J. Ingham, C. Podrini, et al. (2013). "Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes." *Cell* 154.2, pp. 452–64.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilson, A., E. Laurenti, G. Oser, R. C. van der Wath, W. Blanco-Bose, M. Jaworski, S. Offner, C. F. Dunant, L. Eshkind, E. Bockamp, et al. (2008). "Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair". *Cell* 135.6, pp. 1118–1129.

Xiao, T., J. Wallace, and G. Felsenfeld (2011). "Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity." *Molecular and cellular biology* 31.11, pp. 2174–83.

Xu, G., R. Kanezaki, T. Toki, S. Watanabe, Y. Takahashi, K. Terui, I. Kitabayashi, and E. Ito (2006). "Physical association of the patient-specific GATA1 mutants with RUNX1 in acute megakaryoblastic leukemia accompanying Down syndrome". *Leukemia* 20.6, pp. 1002–1008.

Xu, H., K. Balakrishnan, J. Malaterre, M. Beasley, Y. Yan, J. Essers, E. Appeldoorn, J. M. Thomaszewski, M. Vazquez, S. Verschoor, et al. (2010). "Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice". *PLoS ONE* 5.8.

Yan, J., M. Enge, T. Whitington, K. Dave, J. Liu, I. Sur, B. Schmierer, A. Jolma, T. Kivioja, M. Taipale, et al. (2013). "Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites". *Cell* 154.4, pp. 801–813.

Yoshida, K., T. Toki, Y. Okuno, R. Kanezaki, Y. Shiraishi, A. Sato-Otsubo, M. Sanada, M.-j. Park, K. Terui, H. Suzuki, et al. (2013). "The landscape of somatic mutations in Down syndrome-related myeloid disorders." *Nature genetics* 45.11, pp. 1293–9.

Yu, C., A. B. Cantor, H. Yang, C. Browne, R. A. Wells, Y. Fujiwara, and S. H. Orkin (2002). "Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo." *The Journal of experimental medicine* 195.11, pp. 1387–95.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, et al. (2008). "Model-based Analysis of ChIP-Seq (MACS)". *Genome Biology* 9.9, R137.

Zhu, L. J., C. Gazin, N. D. Lawson, H. Pagès, S. M. Lin, D. S. Lapointe, and M. R. Green (2010). "ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data". *BMC Bioinformatics* 11.1, p. 237.

Zuin, J., J. R. Dixon, M. I. J. A. van der Reijden, Z. Ye, P. Kolovos, R. W. W. Brouwer, M. P. C. van de Corput, H. J. G. van de Werken, T. A. Knoch, W. F. J. van IJcken, et al. (2014). "Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells". *Proc Natl Acad Sci USA* 111.3, pp. 996–1001.