Publicly Accessible Penn Dissertations

2013

Dissecting Molecular Pathways That Underlie Disease-Causing Gata1 Mutations

Amy Erin Campbell

University of Pennsylvania, amyerincampbell@gmail.com

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

Part of the [Genetics Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Campbell, Amy Erin, "Dissecting Molecular Pathways That Underlie Disease-Causing Gata1 Mutations" (2013). *Publicly Accessible Penn Dissertations*. 839.
<https://repository.upenn.edu/edissertations/839>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/839>
For more information, please contact repository@pobox.upenn.edu.

Dissecting Molecular Pathways That Underlie Disease-Causing Gata1 Mutations

Abstract

Each mammalian cell type has a unique gene expression pattern that supports its specialized function. Mutations in factors that regulate gene expression can disrupt normal function and cause human disease, though the mechanistic consequences of these defects are often unknown. Here, we address how alterations in the transcription factor GATA1 lead to distinct hematologic disorders by combining structural, biochemical, and genomic approaches with gene complementation systems that examine GATA1 function in biologically relevant cellular contexts. We first investigated missense mutations in the GATA1 N-terminal zinc finger (NF) and found that NF mutations impair association with essential GATA1 cofactors. Several NF mutations diminish FOG1 binding, resulting in greatly reduced transcriptional activation and repression. This severely impairs erythroid and megakaryocyte maturation and correlates with pronounced clinical phenotypes. Notably, clinical severity parallels the degree of FOG1 disruption. Unexpectedly, NF mutations shown to disrupt DNA binding of GATA1 in vitro did not measurably affect target gene occupancy in vivo. Rather, one of these falls into a subset of mutations that diminish TAL1 complex binding. Reduced association with the TAL1 complex moderately impairs transcriptional activation, resulting in subtle defects in erythroid and megakaryocyte development that correlate with relatively mild disease presentations. Remarkably, different substitutions at the same amino acid position can selectively inhibit TAL1 complex or FOG1 binding, producing distinct cellular and clinical phenotypes. We next examined splice site mutations in the second exon of GATA1 that lead to the expression of an amino-truncated protein called GATA1 short (GATA1s). We found that GATA1s was significantly impaired in binding to erythroid-specific target genes, while occupancy at megakaryocyte-specific genes was normal. This results in a strongly diminished erythroid gene expression program and inhibits erythroid maturation, similar to the phenotype observed in patients. In concert, our findings uncover novel molecular mechanisms that link genetic defects in GATA1 to cellular and human phenotypes. Applying this knowledge to the clinic should improve patient diagnosis, classification, and treatment. More broadly, this work highlights the power of gene complementation assays for elucidating the underlying basis of disease, and serves as a model for the study of other disease-causing mutations.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Gerd A. Blobel

Subject Categories

Genetics | Molecular Biology

DISSECTING MOLECULAR PATHWAYS THAT UNDERLIE DISEASE-
CAUSING GATA1 MUTATIONS

Amy E. Campbell

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

Gerd A. Blobel, M.D., Ph.D.
Professor of Pediatrics

Graduate Group Chairperson

Daniel S. Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee

Roger A. Greenberg, M.D., Ph.D., Associate Professor of Cancer Biology

Stephen A. Liebhaber, M.D., Professor of Genetics

John I. Murray, Ph.D., Assistant Professor of Genetics

Nancy A. Speck, Ph.D., Professor of Cell and Developmental Biology

DISSECTING MOLECULAR PATHWAYS THAT UNDERLIE DISEASE-CAUSING GATA1
MUTATIONS

COPYRIGHT

2013

Amy Erin Campbell

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/2.0/>

ACKNOWLEDGMENTS

I would like to thank the following for scientific support:

Blobel Lab

Current

Gerd Blobel
Chris Edwards
Chris Hsiung
Sarah Hsu
Peng Huang
Kristen Jahn
Kiwon Lee
Jeremy Rupon
Aaron Stonestrom
Hongxin Wang
Rena Zheng

Former

Jordan Achtman
Wulan Deng
Greg Gregory
Stephan Kadauke
Harsh Kavi
Janine Lamonica
Annarita Miccio
Jan Pawlicki
Maheshi Udugama
Ariel Yang
Lei Ying

Mackay and Matthews Lab

Joel Mackay
Jacqueline Matthews
Lorna Wilkinson-White

Chou Lab

Stella Chou
Daniel VanDorn

Thesis Committee

Roger Greenberg
Stephen Liebhaber
John Murray
Nancy Speck

Funding

This work was supported by NIH grants
5R37DK058044 and T32DK07780

I would like to thank the following for career development support:

University of Pennsylvania Center for Teaching and Learning
Philadelphia Science Festival
Science Education Academy, Inc.
 Jay Gardner and Danielle Haney
BIOM 555 Course Organizers
 Marisa Bartolomei
 Eric Brown
 Tom Jongsens
Genetics Training Grant T32GM008216
 Meera Sundaram

I would like to thank the following for emotional and moral support:

The Campbell Clan

Rickey Campbell
Julie Campbell
Kirke Campbell
Abbey Campbell
Graeme Campbell

My partner through it all

Dan Denman

The best friends one could ask for

Jayne Taylor Gaubatz
Nate Lavey

ABSTRACT

DISSECTING MOLECULAR PATHWAYS THAT UNDERLIE DISEASE-CAUSING GATA1 MUTATIONS

Amy E. Campbell

Gerd A. Blobel

Each mammalian cell type has a unique gene expression pattern that supports its specialized function. Mutations in factors that regulate gene expression can disrupt normal function and cause human disease, though the mechanistic consequences of these defects are often unknown. Here, we address how alterations in the transcription factor GATA1 lead to distinct hematologic disorders by combining structural, biochemical, and genomic approaches with gene complementation systems that examine GATA1 function in biologically relevant cellular contexts. We first investigated missense mutations in the GATA1 N-terminal zinc finger (NF) and found that NF mutations impair association with essential GATA1 cofactors. Several NF mutations diminish FOG1 binding, resulting in greatly reduced transcriptional activation and repression. This severely impairs erythroid and megakaryocyte maturation and correlates with pronounced clinical phenotypes. Notably, clinical severity parallels the degree of FOG1 disruption. Unexpectedly, NF mutations shown to disrupt DNA binding of GATA1 *in vitro* did not measurably affect target gene occupancy *in vivo*. Rather, one of these falls into a subset of mutations that diminish TAL1 complex binding. Reduced association with the TAL1 complex moderately impairs transcriptional activation, resulting in subtle defects in erythroid and megakaryocyte development that correlate with relatively mild disease presentations. Remarkably, different substitutions at the same amino acid position can selectively inhibit TAL1 complex or FOG1 binding, producing distinct cellular and clinical phenotypes. We next examined splice site mutations in the second exon of GATA1 that lead to the expression of an amino-truncated protein called GATA1 short (GATA1s). We found that GATA1s was significantly impaired in binding to erythroid-specific target genes, while

occupancy at megakaryocyte-specific genes was normal. This results in a strongly diminished erythroid gene expression program and inhibits erythroid maturation, similar to the phenotype observed in patients. In concert, our findings uncover novel molecular mechanisms that link genetic defects in GATA1 to cellular and human phenotypes. Applying this knowledge to the clinic should improve patient diagnosis, classification, and treatment. More broadly, this work highlights the power of gene complementation assays for elucidating the underlying basis of disease, and serves as a model for the study of other disease-causing mutations.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
ABSTRACT	iv
LIST OF TABLES.....	ix
LIST OF FIGURES	x
CHAPTER 1. INTRODUCTION	1
Eukaryotic transcriptional regulation	1
Transcription on DNA.....	2
Transcription on chromatin	4
Transcription in human disease	6
Hematopoiesis	7
Erythrocyte and megakaryocyte development.....	8
Transcription factor GATA1	10
GATA1-associated proteins	12
Human pathophysiologies caused by GATA1 mutations.....	17
CHAPTER 2. EXPERIMENTAL METHODS	23
Constructs.....	23
Cell culture	23
Retroviral infections	24
Western blotting	24
Morphologic analysis	25
Hemoglobin quantification.....	25
Reverse transcription-quantitative polymerase chain reaction	25
Protein production and purification	26
Flow cytometry	26

Transfections and co-immunoprecipitations	26
Isothermal titration calorimetry	27
Chromatin immunoprecipitation	27
Nuclear magnetic resonance spectroscopy	28
Gene expression profiling	29
Primers.....	29
 CHAPTER 3. MISSENSE MUTATIONS IN THE N-TERMINAL ZINC FINGER OF GATA1 IMPAIR BINDING TO ESSENTIAL COFACTORS	 36
Introduction	37
Results	39
Erythroid defects caused by GATA1 NF mutations are recapitulated in an erythroid gene complementation system	39
Megakaryocyte defects caused by GATA1 NF mutations are recapitulated in a megakaryocyte-erythroid gene complementation system	41
A subset of GATA1 NF mutations diminish FOG1 binding	42
GATA1 harboring NF DNA-binding surface mutations occupies target genes normally	43
R216Q and D218G mutations specifically diminish TAL1 complex binding resulting in overlapping gene expression signatures	44
 CHAPTER 4. SPLICE SITE MUTATIONS IN EXON 2 OF GATA1 IMPAIR CHROMATIN OCCUPANCY AT ERYTHROID GENES.....	 80
Introduction	81
Results	82
The N-terminus of GATA1 is required for normal erythroid maturation	82
Differential occupancy of GATA1s at erythroid and megakaryocyte target genes	83
 CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS	 89
A majority of GATA1 NF mutations disrupt FOG1 binding.....	90
If DNA-binding residues of the GATA1 NF aren't binding DNA, what are they doing?	91
A new role for the TAL1 complex in GATA1-mediated disease.....	93
How do different mutations at the same residue selectively disrupt cofactor binding?	94
Aberrant GATA1s expression causes erythroid-specific defects	95

Can individual gene studies be extrapolated genome-wide?	96
Systems for studying disease-causing mutations	97
Do other germline GATA1 mutations exist?.....	100
What can we learn about mutations in other GATA family members?	102
REFERENCES	110

LIST OF TABLES

Table 3.1. Hematologic profiles of patients carrying GATA1 NF mutations.....	74
Table 3.2. Erythroid GATA1 target genes examined by RT-qPCR.....	75
Table 3.3. Megakaryocytic GATA1 target genes examined by RT-qPCR.	76
Table 3.4. FOG1 and DNA binding affinities of GATA1 NF mutants.	77
Table 3.5. GATA1-occupied elements examined by ChIP.....	78
Table 3.6. Summary of disease-associated GATA1 NF mutations.....	79

LIST OF FIGURES

Figure 1.1. Transcriptional regulation.	19
Figure 1.2. Hematopoiesis.	20
Figure 1.3. Transcription factor GATA1.	22
Figure 3.1. Locations of GATA1 NF mutations.	47
Figure 3.2. Expression of GATA1 NF mutants in G1E cells.	48
Figure 3.3. GATA1 NF mutations impair erythroid differentiation.	49
Figure 3.4. Hemoglobin quantification in G1E cells expressing GATA1 NF mutants.	51
Figure 3.5. Impaired transcription in GATA1 NF mutants is not due to delayed maturation.	52
Figure 3.6. GATA1 proteins harboring NF mutations are structured.	53
Figure 3.7. Effect of GATA1 NF mutations on genes encoding heme biosynthetic proteins.	54
Figure 3.8. Expression of GATA1 NF mutants in G1ME cells.	55
Figure 3.9. GATA1 NF mutations impair megakaryocytic maturation.	56
Figure 3.10. Flow cytometry analysis of G1ME cells expressing GATA1 NF mutants.	58
Figure 3.11. Binding of GATA1 NF mutants to FOG1 by co-IP.	59
Figure 3.12. Binding of GATA1 NF mutants to FOG1 by ITC.	60
Figure 3.13. Effect of GATA1 NF mutations on FOG1 chromatin occupancy.	61
Figure 3.14. Effect of GATA1 NF mutations on mast cell genes.	63
Figure 3.15. Binding of GATA1 NF mutants to DNA by ITC.	64
Figure 3.16. Chromatin occupancy of GATA1 NF mutants at palindromic motifs.	65

Figure 3.17. Effect of GATA1 NF mutations on TAL1 complex occupancy.	66
Figure 3.18. Binding of GATA1 NF mutants to LMO2 by NMR spectroscopy.	68
Figure 3.19. Engineered LMO2-disrupting mutations.	69
Figure 3.20. Transcriptional signatures of R216 and D218 mutation pairs.	71
Figure 3.21. Occupancy of the R216W mutant at misregulated genes.	73
Figure 4.1. GATA1s impairs erythroid differentiation.	85
Figure 4.2. Expression of GATA1 and GATA1s in G1ME cells.	86
Figure 4.3. GATA1s chromatin occupancy at erythroid and megakaryocyte genes.	87
Figure 4.4. Effect of GATA1s on erythroid and megakaryocyte gene expression.	88
Figure 5.1. GATA1 NF mutations disrupt cofactor binding: a summary model.	104
Figure 5.2. GATA1s impairs erythropoiesis: a summary model.	106
Figure 5.3. Possible physical causes of differential effects of R216 and D218 mutations.	107
Figure 5.4. Gene set enrichment analysis of G1ME microarray data.	108
Figure 5.5. Absent erythropoiesis from iPSCs harboring GATA1s mutations.	109

CHAPTER 1. INTRODUCTION

Eukaryotic transcriptional regulation

During mammalian development, a single fertilized egg undergoes numerous divisions to generate the trillions of cells that make up a mature organism. Although nearly all of these cells contain the same genetic material, they comprise hundreds of different cell types that carry out highly specialized functions. For example, only pancreatic β -cells produce the peptide hormone insulin to regulate blood glucose levels, while only red blood cells make the oxygen transport protein hemoglobin. This exquisite specialization is achieved by precisely controlling which genes are turned on or off in each cell, thereby establishing a unique gene expression program that determines cellular function and identity.

The human genome contains over 25,000 functional gene products (Djebali et al., 2012), and there are many regulatory mechanisms in place to govern which subset of these genes are expressed in a particular cell at any given time. This regulatory control is exercised at numerous steps, including when and how often a gene is transcribed into RNA, how the RNA is processed, where the RNA is localized, stability of the RNA, RNA translation, and by selectively activating or deactivating a protein once it is made. Normal growth requires this regulation to be spatially and temporally accurate, and defects can lead to aberrant development and disease.

While expression can be regulated at any of the steps mentioned above, for most genes the control of transcription is paramount. Proper transcriptional control ensures that no unnecessary intermediates are synthesized, and therefore best conserves cellular resources. The fundamental mechanisms of transcriptional regulation were elucidated over fifty years ago (Jacob and Monod, 1961), and subsequent studies have refined our understanding and established a well-accepted paradigm (Figure 1.1). In the contemporary model, eukaryotic transcription is a precisely timed event, with regulation occurring at multiple points through the combinatorial and synergistic actions of a host of molecules.

Transcription on DNA

Several DNA regulatory elements must function cooperatively in order to achieve maximal rates of transcription by RNA polymerase II (Pol II). These include: 1) core promoter elements, which are bound by Pol II, 2) promoter proximal elements, which are occupied by proteins that assist the binding of Pol II to the promoter, and 3) distal regulatory elements such as enhancers, silencers, and insulators, which act on the promoter from a considerable distance to modulate expression (Maston et al., 2006). Promoters are typically influenced by multiple distal regulatory elements, and no two genes have the exact same complement of elements (Tjian, 1995). This allows for a combinatorial control that exponentially increases the possible number of unique expression patterns.

The DNA regulatory elements described above contain recognition sites for sequence-specific DNA-binding proteins called transcription factors, which function either to augment or suppress the rate of transcription. Transcription factors can be classified into two groups, general and tissue-specific. General transcription factors (GTFs) are necessary for transcription and include TFIIA, TFIIB, TFIID, TFIIIE, TFIIIF, and TFIIH (Matsui et al., 1980). They assemble onto the core promoter in an ordered fashion to form a preinitiation complex (PIC) that directs Pol II to the transcription start site (Orphanides et al., 1996). However, GTFs alone are only sufficient to direct basal levels of activity; transcription is greatly stimulated or repressed by the presence of tissue-specific transcription factors. These transcription factors accelerate transcription by increasing PIC formation through direct interactions with GTFs, promoting transcription initiation or elongation, or recruiting activities that modify chromatin structure (see below). They can block transcription by inhibiting PIC assembly, impeding or competing for a particular DNA element, or by changing the chromatin state (Lee and Young, 2000). Multiple tissue-specific transcription factors can bind cooperatively to a single regulatory element (Berman et al., 2002), a mechanism that further expands the potential for distinct expression patterns. In an additional layer of regulation, the activities of tissue-specific transcription factors are often modulated by the action

of cofactors. Cofactors do not directly bind DNA, but are instead recruited to regulatory elements via protein-protein interactions with a bound transcription factor where they function to enhance or repress transcriptional activity through the same mechanisms used by transcription factors themselves (Maston et al., 2006).

Once Pol II is recruited to the promoter by transcription factors, the promoter DNA is unwound and transcription initiates. Pol II first produces small, abortive RNAs, but after it attains a stable hold on the DNA and generates a slightly longer RNA molecule, it is able to clear the promoter and engage in transcript elongation (Fuda et al., 2009). At some genes Pol II continues to productively elongate through the gene body to create a full-length transcript. However, at many other loci transcription proceeds only a short distance (20-50 base pairs) before Pol II pauses (Core et al., 2008; Muse et al., 2007; Zeitlinger et al., 2007). Pausing is achieved through the action of DSIF and NELF, which bind and inhibit the polymerase (Wada et al., 1998; Yamaguchi et al., 1999). Pause release and productive elongation occur only once a robust positive signal in the form of the protein P-TEFb arrives. P-TEFb modifies DSIF and NELF to relieve their negative influence, and also modifies Pol II itself (Fujinaga et al., 2004; Marshall et al., 1996; Yamada et al., 2006). This step is increasingly recognized as a major mechanism in gene expression regulation.

As Pol II traverses the gene coding region, factors such as TFIIS, ELL, Elongin, FACT, and PAF regulate its processivity (Bradsher et al., 1993; Kim et al., 2010; Orphanides et al., 1998; Reines et al., 1989; Shilatifard et al., 1996). This elongation machinery additionally interacts with proteins involved in RNA capping and splicing to achieve co-transcriptional processing of the nascent RNA transcript (Montes et al., 2012). When Pol II reaches the end of a gene, it transcribes through a conserved DNA element known as a polyadenylation signal. This first slows and then pauses Pol II, which leads to cleavage and modification of the transcript (Connelly and Manley, 1988; Orozco et al., 2002). This process of transcription termination is mediated by a

number of factors, including CPSF, Cstf, Xrn2, and poly A polymerase (Shandilya and Roberts, 2012).

Transcription on chromatin

Eukaryotic transcription does not occur on a naked DNA template. Rather, genomic DNA is packaged into a proteinacious structure known as chromatin. Chromatin allows the approximately two meters of DNA in each cell to be compacted enough to fit the confines of a nucleus that is $\sim 10 \mu\text{m}$ in diameter. The basic unit of chromatin, the nucleosome, consists of 147 base pairs of DNA wrapped around a protein octamer containing two each of the core histones H2A, H2B, H3, and H4 (Luger et al., 1997). Nucleosomes are connected by stretches of unwrapped linker DNA, which varies in length and is bound by an additional histone, linker histone H1 or H5. This level of compaction results in an 11 nm fiber reminiscent of a “beads-on-a-string” structure (Olins and Olins, 1974; Woodcock et al., 1976). These fibers are then further condensed, with maximal packaging into chromosomes achieved during cell division.

Significant obstacles to transcription arise as a consequence of compacting DNA into chromatin. DNA accessibility is greatly restricted and therefore transcription factors are blocked from binding target sequences, Pol II is prevented from traveling along the gene, and transcription is inhibited. For transcription to occur, chromatin structure must be altered to allow the appropriate machinery access to DNA. The modification of chromatin is carried out by two classes of enzymes – ATP-dependent chromatin remodelers and histone modifiers. By necessitating these additional layers of regulatory complexity, the packaging of DNA into chromatin further expands the potential for unique gene expression patterns and allows for more finely tuned transcriptional outputs.

ATP-dependent chromatin remodelers use the energy of ATP hydrolysis to non-covalently restructure and reposition nucleosomes, thereby altering DNA accessibility to either promote or

repress transcription. The SWI/SNF (mating type switching/sucrose non-fermenting) family of remodelers is generally linked to transcriptional activation, while the ISWI (imitation switch) family typically acts to silence transcription (Saha et al., 2006). In contrast, the CHD (chromodomain helicase DNA-binding) and INO80 (inositol auxotroph 80) families play roles in both activation and repression (Conaway and Conaway, 2009; Murawska and Brehm, 2011). The importance of remodelers is illustrated by the *INF-β* locus, where SWI/SNF slides promoter nucleosomes to activate transcription, which is critical for antiviral immunity (Lomvardas and Thanos, 2001; Zhang et al., 2008).

Histone modifiers covalently alter nucleosomal histones by adding or removing specific chemical groups. The enzymes that attach these groups are referred to as “writers”, and those that remove them are called “erasers”. At least ten distinct classes of modification have been described, including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, and ribosylation (Tan et al., 2011). Most modifications are dynamic, and they adopt predictable genomic distributions that can be used to define specific cellular processes (Rando and Chang, 2009). For example, trimethylation of lysine 36 on histone H3, added by the methyltransferase SETD2 and removed by the demethylase JMJD2, is found in the coding region of actively transcribed genes (Klose et al., 2006; Kouzarides, 2007; Sun et al., 2005).

Histone modifications serve two main functions. First, they can enhance or weaken histone-histone or histone-DNA contacts to directly alter DNA accessibility. In this regard, histone acetylation is thought to directly unfold chromatin by neutralizing positively charged lysine residues that typically contact negatively charged DNA. Indeed, acetylation of lysine 16 on histone H4 impedes 11 nm fiber compaction in vitro (Shogren-Knaak et al., 2006). Second, histone modifications can indirectly alter DNA accessibility by creating interaction surfaces that allow or prevent the docking of effector molecules, called “readers”, which then go on to alter chromatin. “Readers” recognize specific modifications via specialized structural folds such as bromo-, chromo-, and PHD domains (Yun et al., 2011). For example, the PHD domain-containing

protein ING2 binds trimethylated lysine 4 on histone H3 and recruits the Sin3a-HDAC1 deacetylase complex to silence proliferation genes after DNA-damage exposure (Shi et al., 2006). As with remodelers, histone modifications can result in either the activation or repression of gene transcription.

Transcription in human disease

Many clinical phenotypes, from cancer to diabetes to cardiovascular disease, are caused by mutations in transcriptional regulators. Disease-causing mutations can occur within DNA regulatory elements. For example, a single nucleotide change at the *GP1B β* proximal promoter disrupts the binding of a tissue-specific transcription factor, which impairs transcriptional activation and leads to the bleeding disorder Bernard-Soulier Syndrome (Ludlow et al., 1996). More recent analyses are revealing the major role that sequence changes in enhancer elements play in misregulation of gene expression and disease (Maurano et al., 2012). A clear illustration of this phenomenon is seen with dysregulated *SHH* gene expression and limb malformation, where mutations in a *SHH* enhancer cause preaxial polydactyly (Lettice et al., 2003).

Disease-causing mutations are also found in the genes encoding transcription factors and their cofactors. The GTF TFIID is mutated in xeroderma pigmentosum (Lehmann, 2001), and numerous tissue-specific transcription factors have been linked to human disorders (Lee and Young, 2013). Examples in this latter category include heart disease caused by *NKX2-5* mutations (Schott et al., 1998), diabetes caused by mutations in *HNF4 α* (Yamagata et al., 1996), and the cancer-predisposing Li-Fraumeni syndrome caused by *TP53* mutations (Varley, 2003). Similarly, alterations to the transcriptional cofactor Mediator have been implicated in cancer (Barbieri et al., 2012) and neurological disorders (Kaufmann et al., 2010; Risheg et al., 2007; Schwartz et al., 2007).

Finally, defects in chromatin-modifying and -binding proteins can lead to disease. For example, the ATP-dependent chromatin remodeler SWI/SNF is mutated in many cancers (Wilson and Roberts, 2011), the methylation “eraser” *UTX* is mutated in acute lymphoblastic leukemia (Mar et al., 2012), and the acetyl “readers” BRD3 and BRD4 are rearranged in NUT midline carcinoma (French et al., 2008).

Given its central place in human disease, modulating transcription for therapeutic benefit is an important goal of biomedical research. Approximately 10% of currently prescribed drugs target transcription factors of the nuclear receptor family (Overington et al., 2006), though non-ligand-binding transcription factors have proven more recalcitrant to targeting and drug development here has been slower. There is a burgeoning effort to develop small molecules against other chromatin proteins (Dawson and Kouzarides, 2012), and early success with bromodomain inhibitors (Dawson et al., 2011; Filippakopoulos et al., 2010; Nicodeme et al., 2010) makes this an exciting area of work. Additionally, the development of engineered zinc finger, transcription activator-like effector (TALE), and RNA-guided CRISPR/Cas DNA-binding proteins has opened the door to targeted activation or repression of individual misregulated gene loci to ameliorate or reverse disease symptoms (Gao et al., 2013; Garriga-Canut et al., 2012; Gilbert et al., 2013; Wilber et al., 2010). Genome editing to correct disease-causing mutations is also possible using these engineered systems (Gaj et al., 2013).

Hematopoiesis

Hematopoiesis is the process by which mature blood cells are formed from the self-renewal and differentiation of pluripotent hematopoietic stem cells (HSCs). It is a complex, multistage event that begins during embryonic development and continues throughout adulthood. In mammals, hematopoiesis occurs in two waves: first in the fetal yolk sac, placenta, and aorta-gonad-mesonephros region (primitive hematopoiesis), and later in the fetal liver and bone marrow

(definitive hematopoiesis), where it continues throughout adulthood. At each of these sites, extrinsic signals instruct HSCs to develop into multipotent progenitors, which become committed precursors that eventually differentiate into mature, specialized cell types (Figure 1.2). These mature cells are short lived, and therefore blood is being constantly regenerated as billions of old cells are replenished with new ones each day (Rieger and Schroeder, 2012).

Erythrocyte and megakaryocyte development

Red blood cells (also referred to as erythroid cells or erythrocytes) and megakaryocytes are closely related hematopoietic cell types. They arise from a common megakaryocyte-erythroid progenitor (MEP) and require some of the same transcriptional regulators for proper development. However, they carry out very different functions. Erythrocytes transport oxygen to all the tissues of the body, while megakaryocytes generate platelets (also referred to as thrombocytes), which play an essential role in blood clotting and wound healing.

When a MEP is exposed to erythroid stimulating signals, it develops into a slowly proliferating, committed erythroid precursor called a BFU-E (burst forming unit-erythroid). BFU-Es further differentiate into rapidly dividing CFU-E (colony forming unit-erythroid). In the presence of the cytokine erythropoietin (EPO), CFU-Es undergo terminal erythroid maturation. Now, morphologically distinct cells are produced with each successive mitosis, beginning with proerythroblasts and followed by basophilic, polychromatic, and orthochromatic erythroblasts. Orthochromatic erythroblasts expel their nuclei to generate reticulocytes, which enter the circulation and finally become erythrocytes. This ordered maturation is accompanied by drastic cellular changes, including nuclear condensation, decreased cell size, hemoglobinization, and changes in membrane organization. The resulting erythrocytes are anucleate $\sim 7 \mu\text{m}$ diameter biconcave disks chock-full of hemoglobin. Their small size and distinctive shape allows them to squeeze through microcapillaries and creates a large surface area for efficient gas exchange.

After ~100 days in circulation, aging erythrocytes are phagocytosed, thus removing old and defective cells (Dzierzak and Philipsen, 2013; Hattangadi et al., 2011).

In contrast, when a MEP is exposed to megakaryocyte stimulating signals, it develops into a committed megakaryocytic precursor called a BFU-Meg (burst forming unit-megakaryocyte). BFU-Meg cells give rise to CFU-Meg (colony forming unit-megakaryocyte), which undergo terminal maturation to produce promegakaryoblasts, megakaryoblasts, promegakaryocytes, and megakaryocytes. Unlike in erythroid development, terminal megakaryocyte maturation does not involve cell division. Instead, distinct cell stages arise through endomitosis and cytoplasmic maturation, which increase DNA content and cell size. By undergoing repeated cycles of DNA replication without division, a diploid ($2n$) promegakaryoblast becomes a $256n$ megakaryocyte. (Most megakaryocytes undergo three endomitoses and are $16n$.) Concurrently, the cytoplasm grows and fills with specialized organelles and invaginated membrane. The resulting megakaryocyte is big (~50-100 μm diameter) and contains a single, large, multilobulated, polyploid nucleus surrounded by a complex cytoplasmic network. This network is then divided up and packaged into platelets via the formation of long, thin filament extensions called proplatelets. After platelet release, the megakaryocyte (composed mostly of the remaining nucleus) undergoes apoptosis and is consumed by macrophages (Deutsch and Tomer, 2006; Pang et al., 2005).

Erythrocyte and megakaryocyte development are regulated by a relatively small number of tissue-specific transcription factors and cofactors. These factors function to activate lineage-specific genes and repress genes associated with immature, proliferative states and alternative cell fates. Some factors, like GATA1 and FOG1, are required for the maturation of both cell types. Others are lineage-specific. For example, EKLF is important for erythroid but not megakaryocyte development, while the opposite is true for Fli1 (Orkin and Zon, 2008). However, as of yet, no single unique factor has been identified that determines lineage choice of the MEP.

Transcription factor GATA1

The transcription factor GATA1 was discovered in the late 1980s as researchers explored mechanisms that control hemoglobin synthesis (Evans and Felsenfeld, 1989; Evans et al., 1988; Tsai et al., 1989; Wall et al., 1988). Soon after, loss-of-function studies established GATA1 as absolutely critical for both erythrocyte and megakaryocyte development. Whole-body deletion of GATA1 in mice leads to death around embryonic day 11.5 due to a lack of mature erythroid cells (Fujiwara et al., 1996). A block in differentiation and apoptosis at the proerythroblast stage account for the failure to produce mature red blood cells (Pevny et al., 1995; Weiss and Orkin, 1995; Weiss et al., 1994). The early death of GATA1-null embryos precludes direct study of their megakaryocytes. However, tissue-specific ablation of GATA1 in murine megakaryocytes markedly decreases platelet production by causing maturation arrest and hyperproliferation of immature cells (Shivdasani et al., 1997; Vyas et al., 1999).

In addition to insights gleaned from animals, much has been learned about the role of GATA1 in erythroid and megakaryocyte development through the use of two cell lines, GATA1-null erythroid (G1E) and GATA1-null megakaryocyte-erythroid (G1ME). Both G1E and G1ME cells were derived from GATA1-null murine embryonic stem cells. G1E cells proliferate as proerythroblasts and undergo terminal erythroid differentiation upon restoration of GATA1 function (Weiss et al., 1997). A G1E-ER subclone of these cells was engineered to stably express a conditional form of GATA1 (GATA1 fused to the ligand-binding domain of the estrogen receptor, GATA1-ER) that imparts estradiol (E2)-dependent erythroid maturation in a manner largely reproducing that of wild-type cells (Pilon et al., 2011; Tsang et al., 1997; Welch et al., 2004). G1ME cells self-renew in culture as bipotential megakaryocyte-erythroid progenitors and undergo terminal differentiation towards the erythroid and megakaryocyte lineages when reconstituted with GATA1 and grown in the appropriate cytokine environment (Stachura et al., 2006). These easily manipulatable systems enable the study of GATA1 in its natural context at physiological expression levels and provide a homogenous cellular and genetic background that reveals subtle

defects. Further, large amounts of material can be readily obtained for biochemical studies. Using these gene complementation systems, valuable insights have been gained regarding structure/function relationships (Weiss et al., 1997), post-translational modifications (Hung et al., 1999; Lee et al., 2009), novel target genes (Kihm et al., 2002; Shirihaei et al., 2000), mechanisms of cell cycle regulation (Rylski et al., 2003), changes in chromatin structure (Jing et al., 2008; Letting et al., 2003; Vakoc et al., 2005), and lineage choice (Chou et al., 2009).

GATA1 promotes the maturation of erythrocytes and megakaryocytes by activating the transcription of lineage-specific genes and silencing the transcription of genes associated with immature, proliferative states and alternative cell fates (Ferreira et al., 2005). This task is accomplished through four functional domains: two highly conserved zinc finger domains of the C4 (Cys-X₂-Cys-X₁₇-Cys-X₂-Cys) type and two activation domains (Figure 1.3). The C-terminal zinc finger (CF) is the primary mediator of DNA binding, recognizing the canonical GATA1 consensus sequence (A/T)GATA(A/G) (Evans et al., 1988; Martin and Orkin, 1990). The N-terminal zinc finger (NF) stabilizes DNA binding at canonical motifs and is required for GATA1 to bind non-canonical GATC and palindromic ATC(A/T)GATA(A/G) motifs in vitro (Martin and Orkin, 1990; Newton et al., 2001; Trainor et al., 1996). The surface of GATA1 that contacts DNA has been defined structurally (Omichinski et al., 1993). In addition to binding DNA, the zinc finger domains mediate protein-protein interactions (discussed below). In particular, the NF directly binds to the essential cofactors FOG1 and LMO2. The N- and C-terminal activation domains (the NAD and CAD) also mediate protein interactions (discussed below). Defined by their ability to confer transcriptional activation in reporter assays in heterologous cells (Martin and Orkin, 1990), the NAD and CAD play important, if poorly understood, roles in vivo (Kaneko et al., 2012; Kuhl et al., 2005; Li et al., 2005; Muntean and Crispino, 2005; Shimizu et al., 2001; 2009) likely by recruiting specific coregulators to GATA1 target genes.

GATA1-associated proteins

GATA1 interacts with a number of other proteins, both tissue-specific and ubiquitously expressed, which modulate its activity and help ensure lineage-specific gene expression. Establishing which of these interactions are functionally significant, determining the mechanisms through which they modulate GATA1 activity, and understanding how they relate to one another is an ongoing challenge.

FOG1

Friend of GATA1 (FOG1) is a hematopoietic-specific protein that was first identified as a GATA1 binding partner by yeast two-hybrid screen (Tsang et al., 1997). Although FOG1 contains nine zinc finger domains, it does not bind DNA. Instead, FOG1 interacts directly with the GATA1 NF via its first, fifth, sixth, and ninth zinc fingers (Fox et al., 1998; 1999; Tsang et al., 1997). The surface of the GATA1 NF that contacts FOG1 has been defined structurally, and is distinct from the surface that binds DNA (Liew et al., 2005). Like GATA1, FOG1 is required for normal erythroid and megakaryocyte differentiation (Tsang et al., 1997; 1998). Disrupting the GATA1-FOG1 interaction impairs normal maturation of these lineages (Chang et al., 2002; Crispino et al., 1999), interrupts activation and repression of most GATA1-regulated genes (Johnson et al., 2007), and results in aberrant activation of mast cell-specific genes (Chlon et al., 2012; Gregory et al., 2010). Importantly, the clinical relevance of the GATA1-FOG1 interaction was made clear by the identification of patients with anemia and thrombocytopenia who have GATA1 mutations that impair FOG1 binding (Nichols et al., 2000) (discussed below). In addition to the functions described above, FOG1 modulates GATA1 chromatin occupancy at a subset of genomic sites (Chlon et al., 2012; Letting et al., 2004; Pal et al., 2004) and is involved in GATA1-dependent chromatin looping (Jing et al., 2008). FOG1 associates with the transcriptional repressors CtBP and NuRD (Fox et al., 1999; Hong et al., 2005), interactions thought to explain GATA1-dependent gene repression. However, although shown to be functionally important, NuRD is present at both

active and repressed genes (Miccio et al., 2010), leaving unresolved the exact mechanisms that determine the repressive activity of GATA1.

TAL1/LMO2/Ldb1/E2A complex

The TAL1 complex, composed of T-cell acute lymphocytic leukemia 1 (TAL1), LIM domain only 2 (LMO2), LIM domain-binding 1 (Ldb1), and E2A, interacts via LMO2 with the GATA1 NF (Osada et al., 1995; Wadman et al., 1997; Wilkinson-White et al., 2011). LMO2 likely functions as a scaffold that bridges the DNA-binding GATA1 and TAL1 proteins (Wadman et al., 1997; Wilkinson-White et al., 2011). The region of the GATA1 NF that contacts LMO2 has been defined structurally, and is distinct from the DNA- and FOG1-binding surfaces (Wilkinson-White et al., 2011). TAL1, LMO2, and Ldb1 are all essential for erythrocyte and megakaryocyte differentiation (Hall et al., 2003; 2005; Li et al., 2010; Mikkola et al., 2003; Porcher et al., 1996; Yamada et al., 1998). Unlike FOG1, association of the TAL1 complex with GATA1-regulated genes favors activated over repressed targets (Cheng et al., 2009; Kassouf et al., 2010; Soler et al., 2010; Tripic et al., 2009; Yu et al., 2009); however, like FOG1, the Ldb1 component of the complex promotes GATA1-mediated chromatin looping (Deng et al., 2012; Song et al., 2007). This complex has been shown to associate with several additional proteins such as the corepressor ETO-2 and the single-stranded DNA-binding protein SSBP2 (Goardon et al., 2006; Meier et al., 2006; Schuh et al., 2005; Xu et al., 2007), which potentially modulate the activity of the complex.

EKLF

Erythroid Krüppel-like factor (EKLF) is an essential, erythroid-specific transcription factor that binds to the zinc finger domains of GATA1 (Merika and Orkin, 1995; Miller and Bieker, 1993; Nuez et al., 1995; Perkins et al., 1995; Southwood et al., 1996). EKLF and GATA1 co-occupy erythroid regulatory elements (Tallack et al., 2010), suggesting they cooperate to control red blood cell differentiation. Several developmental steps rely particularly upon EKLF, including

lineage commitment at the MEP stage, the developmental switch between fetal and adult hemoglobin expression, and cell cycle exit during terminal maturation (Frontelo et al., 2007; Pilon et al., 2008; Siatecka and Bieker, 2011; Zhou et al., 2010).

Fli1

Friend leukemia integration 1 (Fli1) is a transcription factor that binds to GATA1 via the zinc finger domains and plays a critical role in megakaryocyte development (Eisbacher et al., 2003; Hart et al., 2000; Kawada et al., 2001). The regulatory elements of many megakaryocyte-specific genes contain occupied, tandem binding sites for GATA1 and Fli1 (Lemarchandel et al., 1993; Tijssen et al., 2011; Wang et al., 2002), suggesting a functional interaction between the two factors. Importantly, Fli1 actively inhibits erythroid differentiation (Athanasίου et al., 2000). Current models suggest that Fli1 and EKLF (see above) play antagonizing roles in MEP lineage choice. While Fli1 represses erythropoiesis and promotes megakaryopoiesis, EKLF acts to promote erythroid development and inhibit the megakaryocytic program (Siatecka and Bieker, 2011).

Gfi-1b

Growth factor independent 1b (Gfi-1b) is a hematopoietic transcription factor essential for megakaryocyte and erythroid development (Saleque et al., 2002). It binds to the CF of GATA1 (Huang et al., 2005; Rodriguez et al., 2005) and typically acts to repress gene expression (Grimes et al., 1996). Gfi-1b is thought to function in regulating cell proliferation during terminal maturation (Duan and Horwitz, 2003).

PU.1

PU.1, a hematopoietic transcription factor essential for the development of the monocytic, granulocytic, and lymphoid lineages, binds to the GATA1 CF (Hromas et al., 1993; McKercher et al., 1996; Nerlov et al., 2000; Rekhtman et al., 1999; Scott et al., 1994; Zhang et al., 1999). In

multipotent common myeloid progenitors, PU.1 and GATA1 are expressed together and act in opposition to balance lineage commitment decisions. Their physical interaction is mutually antagonistic, with GATA1 inhibiting PU.1 by preventing its interaction with the essential cofactor c-Jun (Zhang et al., 1999), and PU.1 blocking GATA1 function by inhibiting its ability to bind DNA (Zhang et al., 2000) and/or by recruiting a repressive complex (Rekhtman et al., 2003; Stopka et al., 2005). Downregulation of PU.1 expression is one of the first events associated with the restriction of differentiation to erythroid and megakaryocyte lineages (Nutt et al., 2005).

CBP/p300

The ubiquitously expressed lysine acetyltransferases CREB-binding protein (CBP) and p300 bind to the GATA1 CF and augment transcriptional activity by increasing histone acetylation at GATA1 target genes (Blobel et al., 1998; Forsberg et al., 2000; Letting et al., 2003). In addition to modifying histone proteins, CBP/p300 acetylate GATA1 itself at two lysine-rich regions C-terminal to each zinc finger domain (Boyes et al., 1998; Hung et al., 1999). This acetylation is required for GATA1 chromatin occupancy and therefore erythroid maturation (Hung et al., 1999; Lamonica et al., 2006).

Brd3

Bromodomain-containing protein 3 (Brd3) is a ubiquitously expressed factor that binds to acetylated GATA1 and, through this interaction, is recruited to both activated and repressed GATA1 target genes (Lamonica et al., 2011). Although the function of Brd3 at these genes is unclear, a stable Brd3-GATA1 interaction is crucial for GATA1's association with chromatin and stimulation of erythroid differentiation (Lamonica et al., 2011). Notably, the related proteins Brd2 and Brd4 remain bound to chromatin during mitosis and facilitate postmitotic transcription (Dey et al., 2000; 2003; 2009; Yang et al., 2008; Zhao et al., 2011), implicating them as conveyors of epigenetic information. Additionally, Brd4 interacts with several well-characterized transcriptional

regulators, including P-TEFb and Mediator (Jang et al., 2005; Rahman et al., 2011; Wu and Chiang, 2007; Yang et al., 2005). Whether the same is true for Brd3 has yet to be established.

RUNX1

Runt-related transcription factor 1 (RUNX1) is a hematopoietic/vasculogenic-specific protein that binds to both the NAD and CAD of GATA1 (Elagib et al., 2003). RUNX1 first came to light because of its involvement in leukemic chromosomal translocations (Nucifora and Rowley, 1995), but it also plays a role in normal megakaryopoiesis (Heller et al., 2005; Ichikawa et al., 2004; Michaud et al., 2002; Song et al., 1999). GATA1 and RUNX1 cooperate to activate megakaryocyte genes in vitro (Elagib et al., 2003) and co-occupy megakaryocyte regulatory elements in vivo (Tijssen et al., 2011), but little else is known about how they function together to promote megakaryocyte maturation.

RB

The ubiquitously expressed tumor suppressor protein retinoblastoma (RB) binds to the GATA1 NAD (Kadri et al., 2009). Mice deficient for RB are nonviable and show erythroid defects (Jacks et al., 1992; Lee et al., 1992). In vitro systems have further demonstrated a role for RB in erythroid cell cycle exit and terminal differentiation (Clark et al., 2004). This suggests the GATA1 NAD may regulate the capacity of GATA1 to control red blood cell survival and/or proliferation via RB binding. A truncated version of GATA1 lacking the NAD is aberrantly expressed in several hematopoietic disorders (discussed below). Whether loss of the GATA1-RB interaction is important for these phenotypes is so far unexplored.

Human pathophysiologies caused by GATA1 mutations

Given the essential role of GATA1 in erythroid and megakaryocyte development, it is not surprising that defects in GATA1 cause human hematologic disorders. The first patients with a mutation in the GATA1 gene were described in 2000 (Nichols et al., 2000), and more reports of disease-associated GATA1 mutations followed. These genetic alterations can be transmitted through the germline or acquired somatically, and they result in various clinical phenotypes.

Germline GATA1 mutations are found in the NF, at the exon 2 splice site, and in the stop codon. Because the GATA1 gene is located on the X chromosome (Zon et al., 1990), disorders caused by these mutations predominantly affect males. The seven reported NF mutations are of the missense variety and produce distinct forms of anemia and thrombocytopenia despite being similarly located within the GATA1 protein. Notably, clinical severity depends on the site and type of amino acid substitution, and different substitutions at the same position lead to disparate phenotypes. The mechanisms underlying this variation are not fully understood, although disrupted cofactor binding has been implicated (Ciovacco et al., 2008). Non-synonymous substitutions or deletions in the splice site of exon 2 impair production of full-length GATA1 and cause anemia in affected individuals (Hollanda et al., 2006; Sankaran et al., 2012). How the truncated form of GATA1 (called GATA1s or GATA1 short) acts at the molecular level to produce an erythroid-specific defect is unclear. Finally, a missense mutation in the stop codon predicted to result in the expression of an elongated GATA1 protein leads to a rare blood group phenotype (Singleton et al., 2013). Future studies of this naturally occurring mutation should illuminate the function of the poorly understood CAD.

Acquired GATA1 mutations are found associated with almost all cases of transient myeloproliferative disorder (TMD) and acute megakaryoblastic leukemia (AMKL) in children with trisomy 21 (Down syndrome). Arising in utero, these alterations consist of insertions, deletions, splice site changes, nonsense mutations, and missense mutations within exon 2 that lead to the usage of a downstream start codon and production of the truncated GATA1 protein, GATA1s

(Malinge et al., 2009). Analogous germline GATA1 mutations found in anemic patients do not cause leukemia, underscoring the requirement for trisomy 21 (and likely secondary mutations) in these malignant diseases.

The following work investigates molecular mechanisms that underlie several naturally occurring, disease-causing mutations in the hematopoietic transcription factor GATA1. Although many alterations in the GATA1 gene have been uncovered and their clinical phenotypes described, little is understood about how these mutations alter GATA1 function to cause hematologic disorders. By combining powerful gene complementation systems with structural, biochemical, and genomic approaches, we shed new light on pathways that lie at the core of the pathophysiology caused by GATA1 mutations.

Figure 1.1. Transcriptional regulation.

General transcription factors (light green ovals and horseshoe) bind specific sequence motifs in the core promoter, including the B recognition element (BRE), the TATA box (TATA), the initiator element (Inr), and the downstream promoter element (DPE), and recruit RNA polymerase II (darker green blob) to the transcription start site (arrow). Tissue-specific transcription factors (various colors and shapes) bind regulatory elements (RE) in the proximal promoter and in distal enhancers and silencers where they interact with cofactors (red and blue ovals; pale mint diamond) to enhance or repress transcription. Distal elements influence promoters through physical contacts that involve looping of the DNA. These interactions take place on chromatin, where nucleosomes (grey) can be remodeled or modified (colored “lollipops”) to further influence the transcriptional state. The complex communications that occur among these transcriptional regulatory components allow for cell type specific control of gene expression.

Figure 1.2. Hematopoiesis.

The anatomical sites and timing of hematopoiesis are depicted in the upper panel. The hematopoietic differentiation hierarchy is diagramed in the lower panel. HSCs are capable of

either self-renewal (curved arrow) or differentiation into mature blood lineages (straight arrows). Differentiation can occur via long described routes (thick arrows) and/or more recently described routes (thin arrows). AGM, aorta-gonad-mesonephros; dpc, days post conception; wpc, weeks post conception; HSC, hematopoietic stem cell; MPP, multipotent progenitor; LT, long-term repopulating; IT, intermediate-term repopulating; ST, short-term repopulating; LMPP, lymphoid-primed MPP; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor; MEP, megakaryocyte-erythrocyte progenitor; CDP, common dendritic progenitor; MDP, monocyte-dendritic progenitor; NK, natural killer. Adapted from (Orkin and Zon, 2008) and (Rieger and Schroeder, 2012).

Figure 1.3. Transcription factor GATA1.

The GATA1 gene is comprised of six exons, which are normally spliced together to produce GATA1 mRNA. The primary GATA1 open reading frame starts in exon 2, though an alternative downstream start codon is also present in exon 3. Depending on the initiation site utilized, translation of GATA1 mRNA will generate full-length GATA1 protein or a truncated version called GATA1s that lacks the N-terminal activation domain (NAD). When exon 2 splicing is disrupted (as is seen in human diseases caused by GATA1 mutations), abnormal production of only GATA1s mRNA and GATA1s protein follow. Similarly, disease-causing mutations in exon 2 that disrupt normal translation of the GATA1 mRNA lead to exclusive use of the downstream start codon and aberrant production of GATA1s protein. NF, N-terminal zinc finger; CF, C-terminal zinc finger; CAD, C-terminal activation domain.

CHAPTER 2. EXPERIMENTAL METHODS

Constructs

The murine GATA1 coding sequence alone or fused to the ligand-binding domain of the estrogen receptor was cloned into the MSCV-based retroviral vector MIGR1 to generate MIGR1-GATA1 and MIGR1-GATA1-ER. Point mutations V205M, G208S, G208R, R216Q, R216W, D218G, D218Y, R202Q, or R217M were introduced by site-directed mutagenesis (QuickChange, Stratagene). Full-length GATA1 and the GATA1s (GATA1 Δ 1-83) mutant were amplified by PCR with a single HA tag (YPYDVPDYA) at the N-terminus and inserted into MIGR1 to generate MIGR1-HA-GATA1 and MIGR1-HA-GATA1s. Murine FOG1 was cloned into MIGR1 with a single FLAG tag (DYKDDDDK) fused to the C-terminus to generate MIGR1-FOG1-FLAG. The GATA1 NF, GATA1(200-248), was amplified by PCR and inserted into pGEX-2T to generate pGEX-2T-GATA1(200-248). Point mutations V205M, G208S, G208R, R216Q, R216W, D218G, or D218Y were introduced by site-directed mutagenesis (QuickChange, Stratagene).

Cell culture

G1E cells were cultured as described (Weiss et al., 1997). G1E cells were maintained in Iscove's Modified Dulbecco's Medium supplemented with 15% fetal bovine serum (FBS), 2% penicillin-streptomycin, 0.6% kit-ligand (KL) conditioned medium (CM), 4.5×10^{-5} M monothioglycerol, and 2 U/mL EPO. KL CM was prepared from Chinese hamster ovary cells engineered to express murine KL (Genetics Institute, Cambridge, MA). Based on comparison with recombinant KL, approximately 50 ng/mL KL CM was used. Where indicated, G1E cells were treated with 100 nM E2 for 24-72 hours. G1ME cells were cultured as described (Stachura et al., 2006). G1ME cells were maintained in Minimum Essential Medium Alpha supplemented with 20% FBS, 1% penicillin-streptomycin, 1% glutamine, and 1% thrombopoietin (TPO) CM. TPO CM was prepared from cells engineered to express murine TPO (Villeval et al., 1997). Based on comparison with

recombinant TPO, approximately 20 ng/uL TPO CM was used. Murine erythroleukemia (MEL) and HEK-293 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS, 2% penicillin-streptomycin, 1% glutamine, and 1% sodium pyruvate. Plat-E cells were cultured in DMEM supplemented with 10% FBS, 1% penicillin-streptomycin, 1% glutamine, 1% sodium pyruvate, 1 µg/mL puromycin, and 10 µg/mL blasticidin.

Retroviral infections

Retroviral infections were carried out as described (Chou et al., 2009; Tripic et al., 2009). The MSCV-based retroviral vector MIGR1 was used to express wild-type or mutant murine GATA1 in G1E or G1ME cells. Viral particles were generated by transient transfection of Plat-E retrovirus packaging cells (Morita et al., 2000) using Lipofectamine 2000 according to manufacturer's instructions, and viral supernatant collected 48 hours after transfection. For retroviral transduction, 1-4 mL of retroviral supernatant was mixed with $2-4 \times 10^6$ cells in the presence of 8 µg/mL polybrene and 10 mM HEPES in 1 well of a 6-well plate and spun at 3200 rpm for 90 minutes at room temperature. Cells were incubated at 37 °C for 3 hours and then resuspended in 5 mL fresh media. EPO (2 U/mL) was added to the G1ME transductions to support erythromegakaryocytic differentiation.

Western blotting

Nuclear extracts were prepared according to standard methods (Dignam et al., 1983). In brief, cells were resuspended in ice-cold hypotonic buffer A and allowed to swell on ice for 30 minutes. After centrifugation, the nuclear pellet was resuspended in ice-cold high salt buffer C and incubated on ice for 30 minutes. After centrifugation at 13,000 rpm for 5 minutes supernatant fractions containing nuclear proteins were collected. Samples were fractionated through polyacrylamide gels and transferred onto nitrocellulose (Bio-Rad) or Immobilon-FL PVDF

(Millipore) membranes. Blots were probed with antibodies against GATA1 (sc-265; Santa Cruz), FOG1 (sc-9361; Santa Cruz), HA (sc-805; Santa Cruz), or β -actin (A3854, Sigma-Aldrich). Signals were detected with autoradiography using chemiluminescence (Thermo Scientific) by addition of horseradish peroxidase-conjugated anti-rat (A5795; Sigma-Aldrich) or anti-rabbit (1858415; Pierce) antibody, or with an Odyssey Imager after addition of infrared fluorescent-labeled anti-rat (926-68029 or 926-32219; LiCor) or anti-goat (926-68024 or 926-32214; LiCor) antibody. Band intensities were quantified using ImageJ.

Morphologic analysis

For morphologic analysis, cells were centrifuged onto a glass slide and stained with May-Grünwald-Giemsa (MGG) (Sigma-Aldrich), benzidine for hemoglobin (Orkin et al., 1975), or acetylcholinesterase (AChE), a rodent megakaryocyte marker (Jackson, 1973). Light microscopy images were acquired with a Zeiss Axioskop 2 microscope, Zeiss Axiocam camera, and Zeiss AxioVision 4.8 software (Carl Zeiss Microimaging) at room temperature.

Hemoglobin quantification

For hemoglobin assay, $5\text{--}10 \times 10^6$ cells were lysed in 100 μL of Drabkin's reagent (D5941; Sigma-Aldrich) and hemoglobin content was quantified by spectrophotometric measurement of absorbance at 540 nm on a SpectraMax 190 microplate reader. Hemoglobin content was normalized to total protein, which was determined via Bradford assay.

Reverse transcription-quantitative polymerase chain reaction

RNA was extracted with TRIzol (Invitrogen) from $0.1\text{--}5 \times 10^6$ cells and purified by phenol-chloroform phase separation or with RNeasy Mini columns (Qiagen). Reverse transcription

reactions were performed using Superscript II (Invitrogen) or iScript (BioRad). Results were quantified using SYBR Green dye (Applied Biosystems) on an ABI Prism 7900HT or ViiA 7. Data were normalized to *β-actin* or *Gapdh*, both producing similar results. See below for primer sequences.

Protein production and purification

Construction, expression and purification of LMO2_{LIM2}-Ldb1_{LID}, dFOG1-F1, and the GATA1 NF have been previously described (Liew et al., 2000; Ryan et al., 2006; Wilkinson-White et al., 2011). GATA1 NF mutants were expressed and purified using the same strategy as wild-type GATA1, with one-dimensional ¹H nuclear magnetic resonance (NMR) spectroscopy used to confirm protein folding.

Flow cytometry

Flow cytometry was performed using FACSCanto or LSRFortessa flow cytometers (BD Biosciences) and data analyzed with BD FACSDiva software (BD Biosciences) or FlowJo software (TreeStar). Fluorescence-activated cell sorting (FACS) was performed with a FACSDiva or FACS Aria II (BD Biosciences). For G1ME lineage analysis, cells were stained in phosphate buffered saline (PBS) with 2% FBS at room temperature for 30 minutes with the following antibodies: anti-Ter119-APC, anti-CD41-PE (BD Biosciences), and anti-CD42-PE (Emfret Analytics).

Transfections and co-immunoprecipitations

Transient transfections of HEK-293 cells were performed using polyethylenimine (PEI) at a ratio of PEI to DNA of 3:1. Cells were harvested 48 hours after transfection and nuclear extracts

prepared as described above. Immunoprecipitation (IP) was performed with anti-FLAG beads (Sigma-Aldrich). After extensive washing, elution was carried out with 3X FLAG peptide (Sigma-Aldrich) and bound proteins analyzed by western blotting.

Isothermal titration calorimetry

Experiments were performed using a VP-ITC microcalorimeter (MicroCal) at 20 °C in 10 mM sodium phosphate, 10 mM NaCl, 1 mM DTT pH 7.5. For FOG1/GATA1 NF titrations, dFOG1-F1 (200 μ M) was titrated into a solution containing GATA1 (19 μ M), G208S (20 μ M), G208R (18 μ M), R216W (18 μ M), or D218Y (19 μ M) (Liew et al., 2005). For GATA1 NF/DNA titrations, GATA1 (189 μ M), G208R (196 μ M), R216W (187 μ M), or D218Y (215 μ M) were titrated into a solution containing a double-stranded 16 base pair oligonucleotide containing a GATC site (19-30 μ M) (Liew et al., 2005). Data were analyzed using Origin 5.0 (MicroCal).

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as described (Letting et al., 2003). Protein-DNA crosslinking was performed by incubating cells in PBS with formaldehyde at a final concentration of 1% for 10 minutes at room temperature with gentle agitation. For anti-FOG1 ChIP, cells were crosslinked in 1.5 mM EGS for 20 minutes at room temperature before formaldehyde treatment (Zeng et al., 2006). Glycine (0.125 M) was added to quench the reaction. Cells were then collected by centrifugation at 2000 rpm for 10 minutes and washed in PBS. Nuclei were isolated by incubation in cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% Nonidet P-40) for 10 minutes on ice followed by centrifugation at 2500 rpm for 5 minutes. Nuclei were lysed in nuclei lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) and the lysate sonicated with a Misonix Sonicator 3000 or a Bioruptor (Diagenode) to reduce the chromatin fragments to an average size of less than 500 base pairs. Soluble chromatin was diluted with IP

dilution buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.01% SDS) and precleared by addition of 50 µg preimmune IgG and 50 µL Protein A/G Sepharose. An aliquot of precleared chromatin was set aside (input), and the remainder of the chromatin incubated with 5-10 µg of antibody or preimmune IgG prebound to Protein A/G Sepharose. Protein A/G Sepharose pellets were washed once with IP wash buffer 1 (20 mM Tris pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS), twice with high salt buffer (20 mM Tris pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.01% SDS), once with IP wash buffer 2 (10 mM Tris pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% Nonidet P-40, 1% sodium deoxycholate), and twice with TE (10 mM Tris pH 8.0, 1 mM EDTA). Immune complexes were eluted twice with 100 µL of IP elution buffer (0.1 M NaHCO₃, 1% SDS). RNaseA (0.02 mg/mL), NaCl (0.3 M), and Proteinase K (0.06 mg/mL) were added and crosslinks reversed by incubation at 65 °C overnight. DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. Purified DNA was resuspended in 60 µL (immunoprecipitated DNA) or 133.4 µL (input DNA) water. DNA was quantified by real-time quantitative PCR with SYBR Green dye (Applied Biosystems) on an ABI Prism 7900HT or ViiA7. Results were normalized to an input standard curve dilution series to correct for any differences in starting cell number or primer amplification efficiency. Antibodies used for IP were GATA1 (sc-265; Santa Cruz), FOG1 (sc-9361; Santa Cruz), LMO2 (AF2726; R&D Systems), TAL1 (sc-12984; Santa Cruz), and HA (sc-805; Santa Cruz). See below for primer sequences.

Nuclear magnetic resonance spectroscopy

Chemical shift perturbation experiments were carried out using ¹⁵N-LMO2_{LIM2}-Ldb1_{LID} (200 µM) in 20 mM Mes, 100 mM NaCl, 1 mM DTT pH 6.5. Increasing amounts of unlabelled GATA1 NF (wild-type or mutant) were added up to one molar equivalent, with ¹H, ¹⁵N HSQC spectra recorded after each addition. The weighted average chemical shift changes for the peaks E118, G132, and V131 were calculated using the equation $\Delta_{avg} = (((\delta H)^2 + (\delta N \times 0.154)^2)/2)^{1/2}$ (Ayed et al., 2001).

Gene expression profiling

TRIzol (Invitrogen) extracted RNA was purified with RNeasy Mini columns (Qiagen). Hybridization (GeneChip Mouse Gene 1.0 ST Array; Affymetrix) and data analysis were performed at the University of Pennsylvania Microarray Core Facility. Data were imported into the Partek Genomics Suite (version 6.6; Partek, Inc.) and normalized using RMA to yield log2-transformed expression intensities for all transcript clusters on the array. After excluding control probe sets, principal components analysis and hierarchical clustering were performed to assess sample relatedness and contrast the global variation between and within replicate groups. One-way analysis of variance across the five conditions (each with three replicates) and all ten possible pairwise comparisons was performed, yielding a P value and fold change for all genes. P values were corrected for false discovery rate by the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995). Genes with fold change ≥ 1.5 and $P < 0.05$ were considered differentially expressed. Additionally, Gene Ontology analysis using Database for Annotation, Visualization, and Integrated Discovery (Dennis et al., 2003) was performed on select gene lists and Gene Set Enrichment Analysis (Mootha et al., 2003; Subramanian et al., 2005) was performed on whole datasets. Microarray data are deposited at Gene Expression Omnibus under accession number GSE43356.

Primers

Mutagenesis primers (5' to 3')

GATA1_V205M

Forward GGCCAGAGAGTGTATGAACTGTGGAGCAACG

Reverse CGTTGCTCCACAGTTCATACACTCTCTGGCC

GATA1_G208S

Forward GAGAGTGTGTGAACTGTTCAGCAACGGCTACTCC

Reverse GGAGTAGCCGTTGCTGAACAGTTCACACACTCTC

GATA1_G208R

Forward GAGAGTGTGTGAACTGTAGAGCAACGGCTACTCC

Reverse GGAGTAGCCGTTGCTCTACAGTTCACACACTCTC

GATA1_R216Q

Forward GGCTACTCCACTGTGGCAGAGGGACAGGACAGG

Reverse CCTGTCCTGTCCCTCTGCCACAGTGGAGTAGCC

GATA1_R216W

Forward GGCTACTCCACTGTGGTGGAGGGACAGGACAGG

Reverse CCTGTCCTGTCCCTCCACCACAGTGGAGTAGCC

GATA1_D218G

Forward CCACTGTGGCGGAGGGGCAGGACAGGTCACTACC

Reverse GGTAGTGACCTGTCCTGCCCCTCCGCCACAGTGG

GATA1_D218Y

Forward CCACTGTGGCGGAGGTACAGGACAGGTCACTACC

Reverse GGTAGTGACCTGTCCTGTACCTCCGCCACAGTGG

GATA1_R202Q

Forward GCCCCTTGTGAGGCCCAAGAGTGTGTGAACTGTGG

Reverse CCACAGTTCACACACTCTTGGGCCTCACAAGGGGC

GATA1_R217M

Forward ACTCCACTGTGGCGGATGGACAGGACAGGTCACTACC

Reverse GGTAGTGACCTGTCCTGTCCATCCGCCACAGTGGAGT

Transcript qPCR primers (5' to 3')

Alad

Forward ACTTTCACCCACTGCTTCGGA

Reverse CGATAGGCTGGACATCATCAGG

Alas2

Forward TATGTGCAGGCCATCAACTACCCA

Reverse TTTCCATCATCTGAGGGCTGTGGT

Alox5

Forward ACTACATCTACCTCAGCCTCATT

Reverse GGTGACATCGTAGGAGTCCAC

Alox12

Forward CACACATGGTGAGGAAATGG

Reverse GATCACTGAAGTGGGGCTGT

Casp3

Forward TGGCAACGGAATTCGAGTCCTTCT

Reverse TGAGCATGGACACAATACACGGGA

Clec4d

Forward CGAGAGTAACGTGCATCCGA

Reverse AACAGGACAGCAGGTCCAAGTAC

Cmtm7

Forward GTGCGACCTGATAATGATCCTC

Reverse GGTGAGCACACGGTAGAAGC

Cpox

Forward ATGCAATTTGAAGCCAGTCCGTGG

Reverse TCCCTTTGTAGAATAGCCTGGCCT

Eraf

Forward GCCATGACAGAATTCCAGCAA

Reverse TTTGGACTTCAGAAAGGTCCTGTAT

Fech

Forward AGAGAAGCGAGGTGGTCATTCTGT

Reverse ATGACTTTGTGGACAGTGGCTCCT

Fyb

Forward CGAAGTTCAACACGGGGAGTA

Reverse TGTATTCCAGAAGGCGAGCTT

Gata2

Forward CACCCCTAAGCAGAGAAGCAA

Reverse TGGCACCACAGTTGACACACT

Gcnt2

Forward CTACGCGGGAAAGTTTTTCGC

Reverse GTAGAGGTTGGGCAGGCTTA

Gp1ba

Forward CTTGTTGCCAACGACCAAGCTGAA

Reverse AAGCCCTTTGGTATTGTGCGAAGC

Gp6

Forward GTCTCTCCCAAGTGACCAGCTT

Reverse GAGCTGAGAGTGAGGGTTTAGCA

Gypa

Forward TCACACGGCCCCTACTGAAGTGT

Reverse TCCCTGCCATCACGCGGAAAAT

Hba-a1

Forward GTGGATCCCGTCAACTTCAAG

Reverse CAAGGTCACCAGCAGGCAGT

Hbb-b1

Forward AACGATGGCCTGAATCACTTG

Reverse AGCCTGAAGTTCTCAGGATCC

Hmbs

Forward ATTCCAAGAGGAGCCCAGCTAG

Reverse CATCATTAAGCTGCCGTGCA

Kit

Forward AGCAGATCTCGGACAGCACC

Reverse TGCAGTTTGCCAAGTTGGAG

Klf1

Forward CACGCACACGGGAGAGAAG
Reverse CGTCAGTTCGTCTGAGCGAG

Lrrc39

Forward AAACACCATGAAGACGTGAAGCGG
Reverse TTCCATCTTCCCTGGTCACCTTCT

Ly75

Forward CACGGACAACCTGCGAATGT
Reverse GTATTGGCTACGGCATATCCATC

Myb

Forward TGACTTTTCGACACATGGCTCCTCA
Reverse AATGCACTTGGTGCTGCTCTCAAC

Ms4a2

Forward TGGTTGGTTTGATATGCCTTTGT
Reverse CACTGCACCCCAGAATGGATA

Pf4

Forward TTCTGGGCCTGTTGTTTCTG
Reverse GATCTCCATCGCTTTCTTCG

Ppox

Forward TGCTGTAGAATTCTGGCCTCCGAA
Reverse TCCATCCTGTCCGTGTGCAGATAA

Reep6

Forward GAAACACCCTCACAGCAGCACAAA
Reverse TGCACACCTTGGAGGAGAGATGTT

Ragd

Forward TTTGGGACTCTTTGCTGCCAGTTG
Reverse TCTTATTGCTCTGGCCTGGAGCTT

Slc4a1

Forward TGGAGGCCTGATCCGTGATA
Reverse AGCGCATCGGTGATGTCA

Thbs1

Forward TAGCTGAGGCGGATCAGCAAATCT
Reverse GGGAAGCCAAAGGAGTCCAAATCA

Urod

Forward CCTACATTTCGTGATGTGGCCA
Reverse GGCAAAATGTCCATCCTTAGCA

Uros

Forward ATGGAACGCAGATTCGGAGA
Reverse TGACCTAATGGCCAGTGAACCT

Vwf

Forward TCATCGCTCCAGCCACATTCCATA

Reverse AGCCACGCTCACAGTGGTTATACA

Zfp1

Forward CCTTGCTACCGCAGTCATCA

Reverse ACCAGATCCCGCAGTCTTTG

ChIP qPCR primers (5' to 3')

Abcb10 -6 kb

Forward AGTCTCTATGCCCAGCTTCTTTGG

Reverse AAGACAACGAGAGGAACAGGCAGT

Alad +4 kb

Forward ACTGTGCAGCCTTCCCTCAAGTAA

Reverse TAGCCACAAGATAAGCCAGTGCCA

Alas2 +2 kb

Forward AGGGCAGGACTTTGCCTCTAATCT

Reverse AGATGTCCCAGTTCCTGCAGGTTT

Capn2 +13 kb

Forward TAATGGGAGTTCCCAGCATTT

Reverse GCACAAGAGAGGATGACCTTAT

Eraf prom

Forward TGCCTGCGTCTCGCTTAGT

Reverse GCTGAGCCCGCCTCATC

Ermap +1.7 kb

Forward GGACAGATTCAGGAGGAGAGTA

Reverse CTTTGCACCTCTGAGCTATGAT

Fli1 prom

Forward GCCCAGTTACATTCATGCAC

Reverse TGCAGACTTCAGGAATCAGG

Gata2 -3.9 kb

Forward GAGATGAGCTAATCCCGCTGTA

Reverse AAGGCTGTATTTTTCCAGGCC

Gp1ba prom

Forward TGGTGGCTAGTAGCTGCAAAGTC

Reverse TTATCAGCTCTCTGCACAGCATT

Gypa prom

Forward GCAGTTATGCAGACCTCTAGTT

Reverse CCTCTATCCGTTGACACACATT

Hbb HS2

Forward GGGTGTGTGGCCAGATGTTT

Reverse CACCTTCCCTGTGGACTTCCT

Hbb HS3

Forward CTAGGGACTGAGAGAGGCTGCTT
Reverse ATGGGACCTCTGATAGACACATCT

Hba-a1 -12 kb

Forward AACCTGACTCAAAACAACAAAGTAA
Reverse GGTTCCTGAGTTTCCTTATCTGCAA

Hbb-b1 prom

Forward CAGGGAGAAATATGCTTGTCATCA
Reverse GTGAGCAGATTGGCCCTTACC

Il9r prom

Forward ACATTGCCGAGGACACAGTTCTCT
Reverse TGCATTGCGGAAGGTGAGTCTGTA

Itga2b prom

Forward TCCTGCTCTTGAATGCTGTG
Reverse GGGAGGAAGTGGGTAAATGTC

Kit +33 kb

Forward TGGCAGTCCTGGTTGTAGCA
Reverse GCTGCAAGCATGCGATCA

Klf1 prom

Forward TCTGCTCAAGGAGGAACAGAGCTA
Reverse GGCTCCCTTTCAGGCATTATCAGA

Lrrc39 prom

Forward TTCCCTGGTGTCTGTAGGAACACA
Reverse GGGCTTCTGTGCAAAGGTTCAACT

Lyl1 prom

Forward TCAGCATTGCTTCTTATCAGCC
Reverse CGCAGAGGCCAGAGGATG

Myh9 +5 kb

Forward CACGATTACGGTGACCTTTCTA
Reverse CTTGACTGTGCAGAAGGAAATG

Pf4 prom

Forward GCTGCTGGCCTGCACTTAAG
Reverse GCCACTGGACCCAAAGATAAAG

Pkhd1l1 +0.4 kb

Forward CCGTTCTTCTTGCTCTCCTTGTGT
Reverse AGCTTACCCTGGAAGTGACAGACA

Rragd +4 kb

Forward CTTGGAATCCGAGGAAATGA
Reverse TGGATATCCTCTGGGGAGTG

Slc4a1 +1.6 kb

Forward ATCAGAAGCAACCTAGAGTCCAGC
Reverse TAAGAGTGTAGGACCAGCAGGCAA

Slc4a1 prom

Forward CTGAGCAGTCAAGCCTTAGTTCAC

Reverse CCTGTCCAGTCCCTAAGGTCTTT

Slc22a4 +1.2 kb

Forward CAGCAATGGTGGGAAGGCAGATTT

Reverse ACAAGCACTGTTTCTGGCAAGAGC

Spna1 prom

Forward ATGCCTCACTTTGTCCTGAGCTGT

Reverse TGTCCCTCTGGGCCTTGTTTATCA

Src +5 kb

Forward TTTCCTGTCCTGAAGTGGGTGGAA

Reverse TGGATGGCTACAGCCACCTTAACT

Thbs1 -45 kb

Forward TCACGCTGTGTTGATGAGAGCAGA

Reverse ACTGGGTAGCAGTTCCAAGGGATT

Tph1 prom

Forward ACACGCTTTCACAGAAACCACCAC

Reverse TCTCCCATTAACCGCGTTTCCTCT

Tubb1 +3 kb

Forward CTGTGTTGACTTGAAGGCCTTTGG

Reverse TGA CTCTGTGGCACATAAGGGTA

Uros +6 kb

Forward TACTGCCTGGAAAGGCAAGTGAGA

Reverse ATCTGCCACTGGTATGTCCCAGAA

Uros +11 kb

Forward ACTGCCTTGGTGCTCGTGTGATAA

Reverse AGCTTGCAGAACTTAGCAGCTTCC

Vwf -11 kb

Forward ATATCAGGCCTTTCCTCCAAGGGT

Reverse GCAACTGCCTGCCATGCTATCAAT

Zfp1 +2 kb

Forward CTTTTCTCCTGCCCAGTCG

Reverse TGCTGTTGCCTCGAACC

CHAPTER 3. MISSENSE MUTATIONS IN THE N-TERMINAL ZINC FINGER OF GATA1 IMPAIR BINDING TO ESSENTIAL COFACTORS

Research presented in this chapter was carried out in collaboration with the laboratory of Joel Mackay and Jacqueline Matthews. Protein production and purification, isothermal titration calorimetry, and nuclear magnetic resonance spectroscopy were performed by Lorna Wilkinson-White, a postdoctoral fellow in the Matthews lab.

This work was originally published in Blood. Amy E. Campbell, Lorna Wilkinson-White, Joel P. Mackay, Jacqueline M. Matthews, and Gerd A. Blobel. Analysis of disease-causing GATA1 mutations in murine gene complementation systems. Blood. 2013. 121(26):5218-5227. © The American Society of Hematology.

Introduction

Erythrocyte and megakaryocyte development are under the control of the hematopoietic-specific transcription factor GATA1 (Fujiwara et al., 1996; Shivdasani et al., 1997). GATA1 promotes the differentiation of these lineages by activating all known erythroid- and megakaryocyte-specific genes and silencing genes associated with the immature, proliferative state and alternative lineages (Ferreira et al., 2005).

GATA1 contains two highly conserved zinc finger domains (see Chapter 1), which function to regulate its transcriptional abilities. The CF binds to the canonical DNA sequence motif (A/T)GATA(A/G) while the NF stabilizes DNA interactions by contacting non-canonical GATC and palindromic ATC(A/T)GATA(A/G) motifs (Martin and Orkin, 1990; Newton et al., 2001; Trainor et al., 1996). The NF also binds coregulators, including the multi-zinc finger protein FOG1 (Tsang et al., 1997). Like GATA1, FOG1 is required for erythroid and megakaryocyte development, and disrupting the GATA1-FOG1 interaction impairs maturation of these lineages (Chang et al., 2002; Crispino et al., 1999; Tsang et al., 1998). Activation and repression of most GATA1-regulated genes requires FOG1 (Johnson et al., 2007; Muntean and Crispino, 2005), as does silencing of mast cell-specific genes (Cantor et al., 2008; Chlon et al., 2012; Gregory et al., 2010). FOG1 also modulates GATA1 chromatin occupancy at a subset of genomic sites (Chlon et al., 2012; Letting et al., 2004; Pal et al., 2004). Additionally, the TAL1 complex, composed of TAL1, E2A, LMO2, and Ldb1, interacts via LMO2 with the GATA1 NF (Wadman et al., 1997; Wilkinson-White et al., 2011). TAL1, LMO2, and Ldb1 are essential for erythrocyte and megakaryocyte differentiation (Li et al., 2010; Porcher et al., 1996; Yamada et al., 1998). TAL1 complex recruitment occurs predominantly at GATA1-activated genes and tends to be depleted at sites where GATA1 functions as a repressor (Cheng et al., 2009; Tripic et al., 2009). The interaction surfaces of GATA1 that contact DNA, FOG1, and LMO2 have been defined previously (Liew et al., 2005; Omichinski et al., 1993; Wilkinson-White et al., 2011).

Missense mutations in the GATA1 NF cause distinct forms of congenital anemia and thrombocytopenia. Although similarly located, the seven reported mutations produce a wide spectrum of phenotypes (Balduini et al., 2004; Del Vecchio et al., 2005; Dührsen et al., 2011; Freson et al., 2001; 2002; Kratz et al., 2008; Mehaffey et al., 2001; Nichols et al., 2000; Phillips et al., 2007; Raskind et al., 2000; Tubman et al., 2007; Yu et al., 2002) (Table 3.1). Clinical severity depends on the site and type of substitution, and different substitutions at the same amino acid position produce disparate phenotypes. Broadly, the diseases fall into two categories: severe thrombocytopenia with pronounced anemia (V205M, G208R, D218Y) and moderate thrombocytopenia with minimal or no anemia (G208S, R216Q, R216W, D218G). There is also one case of congenial erythropoietic porphyria (CEP) associated with a R216W substitution. Five mutations lie on defined surfaces: R216Q and R216W sit on the DNA-binding face, while V205M, G208S, and G208R cluster on the FOG1-binding face (Figure 3.1). D218G and D218Y fall outside these surfaces but diminish FOG1 binding in GST-pulldown experiments (Freson et al., 2001; 2002; 2003). Structural and in vitro studies categorized GATA1 NF mutations into two groups, affecting either DNA or FOG1 binding. However, this classification fails to fully explain the degree of phenotypic variation caused by mutations on the same interaction face. For example, both R216Q and R216W are thought to disrupt DNA binding but the latter causes erythroid porphyria while the former does not (Phillips et al., 2007; Yu et al., 2002). Similarly, it is unknown whether the disparate clinical phenotypes caused by different substitutions at residues G208 (Del Vecchio et al., 2005; Mehaffey et al., 2001) and D218 (Freson et al., 2001; 2002) simply disrupt interaction with FOG1 to different extents or affect GATA1 function in qualitatively different ways. Moreover, D218 falls outside the known FOG1-binding surface, raising the possibility that this residue might connect to other GATA1 cofactors. Understanding how GATA1 NF mutations produce human diseases will enhance our understanding of molecular hematopoiesis and refine clinical care by linking prognosis and potential therapies to patient genotypes.

We examined mechanisms by which missense mutations in the NF alter GATA1 function using G1E and G1ME systems (see Chapter 1), which enable the study of GATA1 mutants in

their natural context at physiological expression levels. G1E cells are erythroid precursors that fail to mature owing to a lack of GATA1 (Weiss et al., 1997). Introducing a conditional form of GATA1 (GATA1 fused to the ligand-binding domain of the estrogen receptor [GATA1-ER]) imparts E2-dependent erythroid maturation in a manner largely reproducing that of wild-type cells (Pilon et al., 2011; Welch et al., 2004). G1ME cells are GATA1-null bipotential megakaryocyte-erythroid progenitors that undergo terminal differentiation toward the erythroid and megakaryocyte lineages when reconstituted with GATA1 and grown in the appropriate cytokine environment (Stachura et al., 2006). Using these cell-based systems, complemented by structural, biochemical, and transcriptome analyses, we comprehensively characterize the effect of NF missense mutations on GATA1 function, and uncover novel pathways underlying GATA1-mediated hematologic disorders.

Results

Erythroid defects caused by GATA1 NF mutations are recapitulated in an erythroid gene complementation system

To determine the capacity of GATA1 NF mutants to promote erythroid maturation, wild-type or mutant GATA1-ER fusions were stably introduced into G1E cells. (From here on, GATA1-ER will be referred to simply as GATA1.) All proteins were expressed equivalently and at levels comparable to endogenous GATA1 (Figure 3.2). Upon E2 treatment, wild-type GATA1 induced erythroid maturation as evidenced by morphology and staining with the hemoglobin dye benzidine (Figure 3.3A). In contrast, the V205M, G208R, and D218Y versions of GATA1 were inactive, and G208S displayed only minimal residual activity (Figure 3.3A). The R216Q, R216W, and D218G mutations produced subtle deficiencies detectable upon hemoglobin quantification (Figure 3.3A and Figure 3.4).

GATA1-induced morphologic transitions were reflected in gene expression changes (Figure 3.3B-E). Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis of erythroid GATA1 target genes after 24 hours of E2 treatment showed that the V205M, G208S, G208R, and D218Y mutations reduced transcriptional activation (Figure 3.3B,D) and repression (Figure 3.3C,E) to <20% of wild-type, while R216Q, R216W, and D218G caused less severe defects. Fundamentally similar gene expression patterns were observed at over 30 genes (Table 3.2). Notably, G208R affected transcription more than its paired mutant G208S, and the R216Q and D218G mutations impaired activation more than repression. Gene expression changes were not simply a reflection of delayed maturation because they were also observed at 48 hours of E2 treatment (Figure 3.5). Importantly, all mutant proteins properly regulated a subset of target genes, including *Zfpm1* and *Clec4d* (Figure 3.3B,C), suggesting that missense mutations do not trigger a global misfolding of the NF. To examine this assumption more rigorously, we obtained one-dimensional ¹H NMR spectra of all mutants for which such analyses had not yet been performed. They revealed substantial tertiary structure in all mutant NFs (Figure 3.6 and (Liew et al., 2005)), indicating that the functional consequences of GATA1 NF mutations are not due to global misfolding.

Among GATA1 mutations, R216W is uniquely associated with CEP. Expression analysis of genes encoding heme biosynthetic enzymes revealed a >50% decrease in *Alas2* and *Uros* levels in cells harboring the R216W mutation (Figure 3.7). Although reduced UROS activity (to <5% of wild-type levels) causes CEP, to what extent the observed transcriptional defect could account for this clinical phenotype remains unclear (Balwani and Desnick, 2012).

In summary, the effects of GATA1 NF mutations on cellular morphology, hemoglobin concentrations, and gene expression profiles in G1E cells essentially mimic clinical erythroid phenotypes (see Table 3.6). These findings validate G1E cells as an appropriate and informative model system with which to study disease-causing GATA1 mutations.

Megakaryocyte defects caused by GATA1 NF mutations are recapitulated in a megakaryocyte-erythroid gene complementation system

To test GATA1 NF mutants in a megakaryocytic context, we expressed them in G1ME cells (GATA1-ER fusions are not functional in these cells) and cultured the cells with TPO and EPO to support erythromegakaryocytic differentiation. All proteins were expressed equivalently and at levels comparable to endogenous GATA1 (Figure 3.8). Wild-type GATA1 promoted megakaryocyte maturation, evidenced by large CD42-positive cells with multilobular nuclei that stained positive for acetylcholinesterase (Figure 3.9A and Figure 3.10). The V205M, G208S, G208R, and D218Y mutations caused a marked deficiency in megakaryocyte maturation, while R216Q, R216W, and D218G led to mild defects (Figure 3.9A and Figure 3.10). Importantly, the ability of GATA1 NF mutants to promote the erythroid maturation of G1ME cells was consistent with that observed in G1E cells (compare Figure 3.3 to Figure 3.10).

Megakaryocytic GATA1 target gene expression was measured by RT-qPCR analysis of FACS-purified CD42-positive cells (Figure 3.9B-E). The V205M, G208S, G208R, and D218Y mutations reduced both activation (Figure 3.9B,D) and repression (Figure 3.9C,E) to <20% of that achieved by wild-type GATA1, while R216W had only a mild, gene-specific effect on activation. Although cells harboring the R216Q and D218G mutations displayed no gross morphological deficiencies, there were notable changes in transcriptional activities. Specifically, both mutants displayed a ~60% loss of activation but only a 40% reduction in repression (Figure 3.9D-E). These gene expression changes may explain the thrombocytopenia reported in these patients and suggest that defects caused by R216Q and D218G mutations manifest themselves at a later stage in megakaryocytic maturation or platelet production, which are not detected in this assay. The R216Q and D218G mutations affect transcription in megakaryocytes more strongly than in erythroid cells (compare Figure 3.3D-E to Figure 3.9D-E), suggesting differential sensitivities to diminished GATA1 function between the two cell lineages. This is consistent with the presentation of thrombocytopenia but minimal anemia in these patients. Essentially, the same patterns of

misregulation were seen at over 20 genes (Table 3.3). Importantly, all mutant proteins properly regulated a subset of target genes, including *Reep6* and *Gcnt2* (Figure 3.9B,C), indicating that no mutation completely abrogated GATA1 function in megakaryocytes.

In summary, the ability of mutant GATA1 proteins to promote megakaryocyte maturation in G1ME cells as assessed by morphology, acetylcholinesterase staining, CD42 expression, and transcriptional profiles generally parallels the degree of thrombocytopenia seen in patients (see Table 3.6). In lieu of primary cells from patients, G1E and G1ME cells provide convenient, faithful, and robust systems in which to study GATA1 mutations.

A subset of GATA1 NF mutations diminish FOG1 binding

Except for V205M, studying the impact of GATA1 NF mutations on cofactor binding has been limited to in vitro protein association assays using select FOG1 zinc fingers but not the entire molecule (Freson et al., 2001; 2002; 2003; Mehaffey et al., 2001; Nichols et al., 2000; Yu et al., 2002). We compared the binding of all GATA1 NF mutants to full-length FOG1 by co-immunoprecipitation (co-IP) following expression in HEK-293 cells. The V205M, G208R, and D218Y mutations diminished FOG1 binding by ~70%, 80% and 50%, respectively, while G208S caused only a ~20% reduction (Figure 3.11). Proteins containing R216Q, R216W, or D218G mutations bound indistinguishably from wild-type GATA1 (Figure 3.11). Similar results were obtained when purified GATA1 NF proteins were examined by isothermal titration calorimetry (ITC) (Figure 3.12 and Table 3.4 and (Liew et al., 2005)).

Because cofactor association is influenced by cellular context, we examined the ability of GATA1 NF mutants to recruit FOG1 to target genes in erythroid cells by ChIP. Importantly, the V205M mutation has been reported to diminish GATA1 chromatin occupancy at select sites (Chlon et al., 2012; Letting et al., 2004). We found that V205M, G208S, G208R, and D218Y reduced GATA1 binding at these sites while the remaining mutations were innocuous in this

regard (Figure 3.13A). To assay FOG1 association independently of GATA1 loading, we examined sites where GATA1 chromatin occupancy was FOG1-independent, including single and palindromic motifs near active and repressed genes (Figure 3.13B). Consistent with co-IP data, V205M, G208R, and D218Y severely diminished FOG1 recruitment, G208S had a mild impact, and R216Q, R216W, and D218G had little to no effect (Figure 3.13C-D). Similar observations were made at over 20 target sites (Table 3.5). Furthermore, as previously described (Cantor et al., 2008; Chlon et al., 2012; Gregory et al., 2010), mutations that diminish the GATA1-FOG1 interaction caused aberrant activation of mast cell genes (Figure 3.14).

Two results are especially notable. First, G208S and G208R differ in the degree to which they disrupt binding to FOG1, which mirrors the associated disease severity in patients and supports the notion that inhibition of FOG1 binding, and not another undefined GATA1 cofactor, accounts for the phenotype. Second, substitution of tyrosine for aspartic acid at position 218, which falls outside the known FOG1 interaction domain, inhibits FOG1 binding, while a glycine substitution at the same residue does not (see also below), implicating critical features in the mode of NF interactions that were not predicted by structural studies. In summary, disruption of FOG1 binding in vitro is matched by failure to recruit FOG1 in vivo. This diminishes GATA1 chromatin occupancy at select sites, affects both gene activation and repression, and generally produces the most pronounced defects in erythroid and megakaryocytic differentiation in cell-based assays and in patients.

GATA1 harboring NF DNA-binding surface mutations occupies target genes normally

Because R216 contacts DNA in structural studies (Omichinski et al., 1993), it has been proposed that R216Q and R216W cause disease by disrupting GATA1 DNA binding (Phillips et al., 2007; Yu et al., 2002). We measured by ITC the affinities of wild-type or mutant GATA1 NF proteins for a 16 base pair oligonucleotide containing a GATC site (Figure 3.15 and Table 3.4 and

(Liew et al., 2005)). NF proteins with V205M, G208S, G208R, D218G, or D218Y mutations bound DNA with similar affinities as wild-type. In contrast, the R216Q and R216W mutants showed no measurable binding to DNA. Based on these and previous data (Liew et al., 2005; Newton et al., 2001; Yu et al., 2002), R216Q and R216W would be expected to disrupt GATA1 binding to palindromic and GATC motifs in vivo. Surprisingly, ChIP analysis revealed that GATA1 harboring R216Q or R216W mutations bound all examined target sites normally, including those containing palindromic motifs and regardless of whether strong or weaker GATA1 occupied sites were considered (Figure 3.13A-B and Figure 3.16A). In accordance with in vitro results, V205M, G208S, G208R, D218G, and D218Y mutants showed normal GATA1 chromatin occupancy at all sites except those at which association with FOG1 is required (Figure 3.13A-B and Figure 3.16B). Thus, in contrast to in vitro observations, mutations on the DNA-binding surface of the NF do not significantly impair GATA1 target site occupancy in vivo. Therefore, these mutations likely cause human disease through alternate mechanisms.

R216Q and D218G mutations specifically diminish TAL1 complex binding resulting in overlapping gene expression signatures

Since the R216Q, R216W, and D218G mutations do not measurably disrupt FOG1 recruitment or DNA binding in vivo, we examined whether they affect association with LMO2, a member of the TAL1 complex, because it interacts directly with the GATA1 NF in a manner permitting simultaneous FOG1 binding (Wilkinson-White et al., 2011). Although no NF mutations affect residues known to bind LMO2, R216 and D218 flank R217, which contributes to the GATA1-LMO2 interaction (Wilkinson-White et al., 2011). ChIP analysis revealed that the R216Q and D218G mutations diminished LMO2 occupancy at GATA1 target genes (Figure 3.17A-B). This trend was more apparent when the LMO2 ChIP signals were normalized to those of GATA1 (Figure 3.17C). In contrast, none of the remaining mutations impaired LMO2 recruitment (Figure 3.17A-C). The same binding profiles were observed for TAL1 (Figure 3.17D), suggesting the

entire TAL1 complex is recruited less efficiently by the R216Q or D218G mutants. Similar occupancy profiles were observed at over 20 target sites (Table 3.5). These results were surprising given that R216Q falls on the DNA-binding surface and D218Y disrupts FOG1 binding. However, a shared molecular defect is consistent with the similar clinical phenotypes caused by R216Q and D218G mutations. Furthermore, a qualitative difference in TAL1 complex binding between R216Q and R216W, and between D218G and D218Y, might explain each pair's divergent patient presentations. Notably, disruption of TAL1 complex binding by R216Q and D218G did not reduce GATA1 chromatin occupancy at any site (Figure 3.13A-B and Figure 3.16), suggesting that unlike FOG1, association with the TAL1 complex is not required for stable binding of GATA1 to its target genes. Additionally, disruption of TAL1 complex recruitment impinges on gene activation more so than on repression (Figure 3.3D-E, Figure 3.9D-E, and Figure 3.14), consistent with previous reports of this cofactor predominantly occupying GATA1-activated genes (Cheng et al., 2009; Kassouf et al., 2010; Soler et al., 2010; Tripic et al., 2009; Yu et al., 2009).

Due to the low affinity of the GATA1-LMO2 interaction (Wilkinson-White et al., 2011), co-IP was unreliable for comparing the binding of GATA1 NF mutants to LMO2 (data not shown). Therefore, we turned to two-dimensional ^1H , ^{15}N NMR spectroscopy. The addition of purified GATA1 NF protein containing a R216W or D218Y mutation to a preformed LMO2-Ldb1 complex caused peak shifts similar to those induced by wild-type protein, indicating that these mutations did not affect the interaction between the GATA1 NF and LMO2 (Figure 3.18 and data not shown). Titration of D218G induced chemical shift changes that were smaller in magnitude than those induced by wild-type protein, consistent with a reduction in LMO2 binding affinity (Figure 3.18). However, the signals were qualitatively similar and shifted in the same directions, indicating that the nature of the interaction was fundamentally unaltered compared to wild-type. Addition of R216Q did not result in significant shifts to any peaks in the spectrum, demonstrating that this mutation markedly impairs LMO2 binding (Figure 3.18). Together, these data support our *in vivo* results showing that R216Q and D218G inhibit recruitment of LMO2 to GATA1 bound genes.

To strengthen our finding that the R216Q and D218G mutations affect LMO2 binding, we compared them to two GATA1 substitutions (R202Q and R217M) previously shown to diminish LMO2 binding in vitro (Wilkinson-White et al., 2011). In G1E cells, GATA1 harboring R202Q or R217M induced erythroid maturation and transcriptional activities almost perfectly matching the R216Q and D218G mutants (Figure 3.19A-D, compare to Figure 3.3 and Figure 3.4). ChIP confirmed that R202Q and R217M impaired recruitment of LMO2 but not FOG1 (Figure 3.19E-F). This strongly supports the idea that TAL1 complex disruption by R216Q and D218G mutations accounts for their disease-causing effects, implicating for the first time the TAL1 complex in the pathogenesis of disorders caused by GATA1 mutations.

To obtain a broad and unbiased comparison of wild-type GATA1 and paired mutants that affect TAL1 complex binding (R216Q and D218G) and those that do not (R216W and D218Y), we examined transcriptomes by microarray. Hierarchical clustering analysis revealed that R216Q and D218G were highly similar (Figure 3.20A). Notably, there was substantial overlap among genes affected at least twofold by the R216Q and D218G mutations that exceeded the R216Q and R216W overlap (Figure 3.20B). Similar results were observed when a 1.5-fold cut-off was used. ChIP confirmed diminished LMO2 occupancy at genes downregulated by the R216Q and D218G mutations (Figure 3.20C-D). Thus, gene profiling cemented a shared mechanism of action for the R216Q and D218G mutations, namely disruption of TAL1 complex interaction.

Global expression changes induced by GATA1 harboring R216W are most similar to those of wild-type (Figure 3.20A), suggesting that alterations caused by this mutation are subtle. We note that genes severely misregulated by R216W mutation show no defect in GATA1 occupancy (Figure 3.21), further supporting the conclusion that the R216W mutation impairs functions other than DNA binding.

Figure 3.1. Locations of GATA1 NF mutations.

Space filling model of the GATA1 NF from PDB code 1Y0J with DNA binding residues in red (based on PDB ID code 1GAT), FOG1 binding residues in cyan, and LMO2 interacting residues in blue. The locations of disease-associated mutations are noted. The middle structure has been rotated 120 degrees around a horizontal axis from the leftmost model, and the rightmost structure is rotated a further 80 degrees.

Figure 3.2. Expression of GATA1 NF mutants in G1E cells.

Anti-GATA1 and, as a loading control, anti- β -actin western blots of extracts from G1E cells expressing indicated versions of GATA1 after 24 hours of E2 treatment. Wild-type and mutant proteins were expressed equally (**A**) at levels similar to endogenous GATA1 in MEL cells (**B**). Bands were quantified by densitometry and relative GATA1 levels are noted below each panel. ND = not determined. The black vertical line in (**B**) demarcates an unrelated excised lane.

Figure 3.3. GATA1 NF mutations impair erythroid differentiation.

(A) MGG and benzidine staining of G1E cells expressing wild-type or mutant GATA1 after 72 hours of E2 treatment. The percentage of hemoglobin-positive cells is indicated in the upper right corner of each benzidine panel. Scale bars, 20 μ m (left panels) and 50 μ m (right panels). (B-C)

Expression of GATA1-activated (**B**) and GATA1-repressed (**C**) genes after 24 hours of E2 treatment as determined by RT-qPCR, normalized to *β -actin*, and plotted as fold change from uninfected samples. (**D-E**) Average transcriptional profiles after 24 hours of E2 treatment for all activated (**D**, n = 24) and repressed (**E**, n = 6) genes examined. Note that the reduction in transcriptional repression by R216Q and D218G is not statistically significant. * P <0.05. Error bars denote SEM (n = 3) unless otherwise noted.

Figure 3.4. Hemoglobin quantification in G1E cells expressing GATA1 NF mutants.

Hemoglobin concentrations of indicated cells analyzed after 72 hours of E2 treatment.

Representative images of cell pellets are shown above. Error bars denote SEM (n = 3).

Figure 3.5. Impaired transcription in GATA1 NF mutants is not due to delayed maturation.

Expression of GATA1-activated (**A**) and GATA1-repressed (**B**) genes after 24 hours and 48 hours of E2 treatment as determined by RT-qPCR, normalized to β -actin, and plotted as fold change from uninfected samples. Error bars denote SEM (n = 3).

Figure 3.6. GATA1 proteins harboring NF mutations are structured.

One-dimensional ^1H NMR spectra of indicated GATA1 NF proteins showing a well-defined tertiary fold.

Figure 3.7. Effect of GATA1 NF mutations on genes encoding heme biosynthetic proteins.

(A) Expression of all eight heme biosynthetic enzymes in G1E cells expressing indicated GATA1 versions after 24 hours and 48 hours of E2 treatment as determined by RT-qPCR, normalized to β -actin, and plotted as fold change from uninfected samples. (B) Results from (A) at 24 hours were plotted as fold change from GATA1-expressing cells. (C) Expression of additional genes implicated in heme biosynthesis after 24 hours of E2 treatment as determined by microarray. Data are shown as fold change from wild-type. Error bars denote SEM (n = 3).

Figure 3.8. Expression of GATA1 NF mutants in G1ME cells.

Western blot showing GATA1 expression in G1ME lines 72 hours after infection with wild-type or mutant proteins and compared to endogenous levels in MEL cells. β -actin serves as a loading control. Bands were quantified by densitometry and relative GATA1 levels are noted below each panel.

Figure 3.9. GATA1 NF mutations impair megakaryocytic maturation.

(A) MGG and AChE staining of G1ME cells 72 hours after infection with wild-type or mutant GATA1-expressing vector. Scale bars, 20 μ m. (B-C) Expression of GATA1 activated (C) and repressed (D) genes in FACS-purified CD42-positive megakaryocytes 72 hours after transduction

as determined by RT-qPCR, normalized to *β -actin*, and plotted as fold change from uninfected samples. (**D-E**) Average transcriptional activities of all examined activated (**D**, n = 10) and repressed (**E**, n = 6) genes. * P < 0.05. Error bars denote SEM (n = 3) unless otherwise noted.

Figure 3.10. Flow cytometry analysis of G1ME cells expressing GATA1 NF mutants.

(A) Surface expression of the terminal megakaryocyte marker CD42 and the terminal erythroid marker Ter119 as assessed by flow cytometry in G1ME cells 72 hours after infection with wild-type or mutant GATA1. Dot plots (B) show a representative experiment. Percentages refer to the fraction of GFP-positive cells expressing CD42 or Ter119. Approximately 45% of transduced cells were GFP-positive. Error bars denote SEM (n = 3).

Figure 3.11. Binding of GATA1 NF mutants to FOG1 by co-IP.

(A) Wild-type or mutant GATA1 was co-expressed with FLAG-tagged FOG1 in HEK-293 cells and analyzed by anti-FLAG IP followed by anti-GATA1 or anti-FOG1 western blotting. Input represents 5% of lysate. (B) Quantification of western blot signals. * $P < 0.05$. Error bars denote SEM ($n = 3$).

Figure 3.12. Binding of GATA1 NF mutants to FOG1 by ITC.

ITC data showing titration of the first zinc finger of the *Drosophila* FOG-family protein (dFOG-F1) into wild-type or mutant GATA1 NF proteins.

Figure 3.13. Effect of GATA1 NF mutations on FOG1 chromatin occupancy.

Anti-GATA1 and anti-FOG1 ChIP in G1E cells expressing indicated GATA1 versions after 24 hours of E2 treatment. (A) Anti-GATA1 ChIP at FOG1-dependent binding sites. (B) Anti-GATA1

ChIP at FOG1-independent binding sites that contain single (*Hbb* HS3, *Gata2* -3.9 kb) or palindromic (*Lyl1* prom) motifs and regulate activated (*Hbb*) or repressed (*Gata2*, *Lyl1*) genes. **(C)** Anti-FOG1 ChIP at FOG1-independent binding sites. **(D)** FOG1 ChIP signals were normalized to GATA1 ChIP signals at each FOG1-independent site. Error bars denote SEM (n = 3).

Figure 3.14. Effect of GATA1 NF mutations on mast cell genes.

Expression of mast cell-specific genes after 24 hours of E2 treatment as determined by RT-qPCR, normalized to β -actin, and plotted as fold change from uninfected samples. Error bars denote SEM (n = 3).

Figure 3.15. Binding of GATA1 NF mutants to DNA by ITC.

ITC data showing the titration of wild-type or indicated mutant versions of the GATA1 NF into a 16 base pair oligonucleotide containing a GATC motif.

Figure 3.16. Chromatin occupancy of GATA1 NF mutants at palindromic motifs.

Anti-GATA1 ChIP in G1E cells expressing wild-type or mutant GATA1 after 24 hours of E2 treatment using primers spanning single (*Hbb* HS2) or palindromic (all other) motifs. **(A)** NF mutations that fall on the DNA-binding surface. **(B)** NF mutations that fall outside the DNA-binding surface. Note that the reduced occupancy of V205M, G208S, G208R, and D218Y at *Slc4a1* +1.6 kb, *Slc22a4* +1.2 kb, and *Alad* +4 kb is because GATA1 binding at these sites is FOG1-dependent. Error bars denote SEM (n = 3).

Figure 3.17. Effect of GATA1 NF mutations on TAL1 complex occupancy.

(A-D) Anti-GATA1, anti-LMO2, or anti-TAL1 ChIP in G1E cells expressing wild-type or mutant GATA1 after 24 hours of E2 treatment using primers as in Figure 3.13. For panels (C-D), LMO2

and TAL1 ChIP signals were normalized to GATA1 ChIP signals at each site. Error bars denote SEM (n = 3 for GATA1 and LMO2 ChIP, n = 2 for TAL1 ChIP).

Figure 3.18. Binding of GATA1 NF mutants to LMO2 by NMR spectroscopy.

(A) A portion of the ^1H , ^{15}N HSQC spectra of ^{15}N -LMO2_{LIM2}-Ldb1_{LID} (red peaks) following addition of one equivalent of either GATA1 NF (green), R216Q (cyan), R216W (gold), or D218G (purple).

(B) Relative weighted average change in chemical shift position of resonances from ^{15}N -LMO2_{LIM2}-Ldb1_{LID} following addition of wild-type or mutant GATA1 NF. Shown are the average shifts of three separate resonances for each series of titrations. * P < 0.05. Error bars denote SD.

Figure 3.19. Engineered LMO2-disrupting mutations.

(A) Western blot showing GATA1 levels in G1E cells expressing wild-type or mutant GATA1 after 24 hours of E2 treatment. β -actin serves as a loading control. Bands were quantified by densitometry and relative GATA1 levels are noted below each panel. ND = not determined. (B) MGG and benzidine staining after 72 hours of E2 treatment. The percentage of hemoglobin-positive cells is indicated in the upper right corner of each benzidine panel. Scale bars, 20 μ m (left panels) and 50 μ m (right panels). Compare to R216Q and D218G mutants in Figure 3.3. (C)

Expression of GATA1-regulated genes after 24 hours of E2 treatment as determined by RT-qPCR, normalized to *β-actin*, and plotted as fold change from GATA1-expressing cells. **(D)** Hemoglobin concentrations and representative images of cell pellets after 72 hours of E2 treatment. Compare to R216Q and D218G mutants in Figure 3.4. **(E-F)** Anti-GATA1, anti-FOG1, and anti-LMO2 ChIP after 24 hours of E2 treatment. FOG1 and LMO2 ChIP signals were normalized to GATA1 ChIP signals at this binding site. Error bars denote SEM (n = 3).

Figure 3.20. Transcriptional signatures of R216 and D218 mutation pairs.

(A) Unsupervised hierarchical clustering of G1E cells expressing wild-type or mutant GATA1 based on expression profiling with microarrays. Note that one D218G replicate is indistinguishable from the R216Q replicates. (B) Venn diagrams of direct GATA1 target genes significantly downregulated when compared with wild-type GATA1. (C) Expression of GATA1-regulated genes highly sensitive to TAL1 complex disruption was validated by RT-qPCR, normalized to β -actin, and plotted as fold change from uninfected samples. (D) Anti-GATA1 and anti-LMO2 ChIP in G1E cells expressing wild-type or mutant GATA1 after 24 hours of E2 treatment using primers against GATA1-occupied elements in genes significantly impaired in

response to R216Q and D218G mutations. LMO2 ChIP signals were normalized to GATA1 ChIP signals at each site. Error bars denote SEM (n = 3).

Figure 3.21. Occupancy of the R216W mutant at misregulated genes.

Anti-GATA1 ChIP after 24 hours of E2 treatment using primers recognizing GATA1 binding sites at genes whose activities are at least twofold increased or decreased when compared to wild-type GATA1-expressing cells. Error bars denote SEM (n = 3).

Table 3.1. Hematologic profiles of patients carrying GATA1 NF mutations.

Mutation	Hgb (g/dL) (10.5 - 16)	Hct (%) (33 - 46)	RBC (x10 ⁶ /μL) (3.9 - 5.2)	MCV (μm ³) (70 - 100)	MCH (pg) (23 - 34)	Retic (%) (0.5 - 1.5)	nRBC (%) (0)	RDW (11.5 - 14.5)	Plt (x10 ³ /μL) (150 - 400)	MPV (x10 ⁻¹⁵ L) (8 - 12)	Tpo (pg/mL) (60 - 150)
V205M n = 2	8.4	28.3	3.1	92.4	27.5	3	27	20.6	17.5	-	-
G208S n = 3	-	43	-	100.4	31.9	-	-	-	24	10.5	262.7
G208R n = 3	9.5	30	2.97	104	-	4	-	17.8	13.3	-	1,755
R216Q n = 7	11	42.6	5.4	78	-	-	-	18.3	73.5	9.5	-
R216W n = 1	7.0	-	-	67	<20	5	-	-	70	-	-
D218G n = 4	13.4	42	4.5	-	-	-	-	15.3	34	-	-
D218Y n = 1	9.7	33.5	3.4	-	-	-	-	17.2	12	-	-

Normal values for each parameter are indicated in parentheses. The number of individuals with reported hematologic indices is noted below each mutation. Indices shown are the average of all individuals. Hgb, hemoglobin concentration; Hct, hematocrit; RCB, red blood cell number; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; Retic, reticulocytes as percentage of total red blood cells; nRBC, number of nucleated red blood cells per 100 white blood cells; RDW, red blood cell distribution width; Plt, platelet number; MPV, mean platelet volume; Tpo, thrombopoietin concentration. Data were gathered from Nichols et al., 2000; Mehaffy et al., 2001; Del Vecchio et al., 2005; Kratz et al., 2008; Duhrsen et al., 2011; Raskind et al., 2000; Balduini et al., 2004; Tubman et al., 2007; Phillips et al., 2007; Freson et al., 2001; Freson et al., 2002.

Table 3.2. Erythroid GATA1 target genes examined by RT-qPCR.

Sensitive to GATA1 NF mutations	
<i>Hbb-b1</i>	<i>Gypa</i>
<i>Hba-a1</i>	<i>Hri</i>
<i>Abcb10</i>	<i>Epb4.9</i>
<i>Eraf</i>	<i>Ank1</i>
<i>Alas2</i>	<i>Spna1</i>
<i>Alad</i>	<i>Spnb1</i>
<i>Hmbs</i>	<i>Gstt1</i>
<i>Uros</i>	<i>Cetn4</i>
<i>Urod</i>	<i>Slc4a1*</i>
<i>Cpox</i>	<i>Lrrc39*</i>
<i>Ppox</i>	<i>Rragd*</i>
<i>Fech</i>	<i>Slc22a4*</i>
<i>Myb</i>	<i>Myc</i>
<i>Kit</i>	<i>Hhex</i>
<i>Gata2†</i>	<i>Casp3</i>
Insensitive to GATA1 NF mutations	
<i>Zfp1</i>	<i>Klf1</i>
<i>Lyl1</i>	<i>Adamts5</i>
<i>Clec4d</i>	<i>Rgs18</i>
<i>Tmem44</i>	<i>Car1</i>

*More sensitive to R216Q and D218G

†Less sensitive to G208S

Table 3.3. Megakaryocytic GATA1 target genes examined by RT-qPCR.

Sensitive to GATA1 NF mutations	
<i>Vwf</i>	<i>Pf4</i>
<i>Alox12</i>	<i>Thbs1</i>
<i>Capn2</i>	<i>Itgb3</i>
<i>Ppbp</i>	<i>Gp6</i>
<i>Tubb1</i>	<i>Itga2b</i>
<i>Gata2</i>	<i>Slc22a3</i>
<i>Il1rl1</i>	<i>Fcgr3</i>
<i>Ly75</i>	<i>Cmtm7</i>
Insensitive to GATA1 NF mutations	
<i>Asb1</i>	<i>Pkn3</i>
<i>Reep6</i>	<i>Blvrb</i>
<i>Gcnt2</i>	<i>Clec4e</i>
<i>Kit</i>	<i>Hmga2</i>

Table 3.4. FOG1 and DNA binding affinities of GATA1 NF mutants.

Protein	k_a	
	GATA1:FOG1, M^{-1}	GATA1:DNA, M^{-1}
GATA1	$(4.5 \pm 1.8) \times 10^4^*$	$(1.9 \pm 0.1) \times 10^6^*$
	$(6.6 \pm 0.1) \times 10^4$	$(1.6 \pm 0.3) \times 10^7$
V205M	No binding*	$(2.0 \pm 1.6) \times 10^7^*$
G208S	$(1.3 \pm 0.7) \times 10^{5\dagger}$	$(9.1 \pm 5.2) \times 10^6^*$
G208R	No binding	$(1.8 \pm 0.3) \times 10^7$
R216Q	$(9.4 \pm 0.3) \times 10^4^*$	No binding*
R216W	$(1.7 \pm 0.1) \times 10^5$	No binding
D218G	$(1.9 \pm 0.3) \times 10^4^*$	$>10^7^*$
D218Y	No binding	$(2.4 \pm 1.0) \times 10^7$

Association constants are the average of two independent experiments. The error reports the range associated with that mean.

*Reproduced from Liew et al., 2005.

[†]Previously published data indicated no detectable interaction (Liew et al., 2005). New experiments reveal an interaction with wild-type like affinity but low enthalpy.

Table 3.5. GATA1-occupied elements examined by ChIP.

FOG1-dependent, single GATA motif	
<i>Hbb-b1</i> prom	<i>Hba-a1</i> -12 kb
<i>Slc4a1</i> prom	<i>Rragd</i> +4 kb
FOG1-dependent, palindromic GATA motif	
<i>Slc4a1</i> +1.6 kb	<i>Slc22a4</i> +1.2 kb
<i>Alad</i> +4 kb	
FOG1-independent, single GATA motif	
<i>Hbb</i> HS3	<i>Hbb</i> HS2
<i>Eraf</i> prom	<i>Klf1</i> prom
<i>Gata2</i> -3.9 kb	<i>Lrrc39</i> prom
FOG1-independent, palindromic GATA motif	
<i>Lyl1</i> prom	<i>Zfpm1</i> +2 kb
<i>Spna1</i> prom	<i>Alas2</i> +2 kb
<i>Abcb10</i> -6 kb	<i>Kit</i> +33 kb
<i>Uros</i> +11 kb	<i>Uros</i> +6 kb

Table 3.6. Summary of disease-associated GATA1 NF mutations.

GATA1 mutation	Associated human disease	Erythroid (G1E) maturation	MK (G1ME) maturation	DNA binding		FOG1 binding		TAL1 complex binding		Proposed mechanism of disease
				In vitro	In vivo	In vitro	In vivo	In vitro	In vivo	
Wild-type		+++	+++	+++	+++	+++	+++	+++	+++	
V205M	Severe macrothrombocytopenia Severe dyserythropoietic anemia	—	—	+++	+	—	—	+++	+++	Reduced FOG1 binding
G208S	Macrothrombocytopenia Mild dyserythropoiesis without anemia	+	—	+++	++*	+	+	ND	+++	Reduced FOG1 binding
G208R	Severe macrothrombocytopenia Severe dyserythropoietic anemia	—	—	+++	+	—	—	ND	+++	Reduced FOG1 binding
R216Q	Macrothrombocytopenia Mild dyserythropoiesis without anemia Mild B-thalassemia	++	+	—	+++	+++	+++	—	+	Reduced TAL1 complex binding
R216W	Thrombocytopenia Congenital erythropoietic porphyria	++	++	—	+++	+++	+++	+++	+++	?
D218G	Macrothrombocytopenia Mild dyserythropoiesis without anemia	++	+	+++	+++	+++	+++	+	+	Reduced TAL1 complex binding
D218Y	Severe macrothrombocytopenia Severe dyserythropoietic anemia	—	—	+++	+	—	—	+++	+++	Reduced FOG1 binding

*The reduced DNA binding observed intracellularly is context-dependent and a secondary consequence of impaired FOG1 binding, not a direct effect on the DNA interaction surface. MK, megakaryocyte; ND, not determined.

CHAPTER 4. SPLICE SITE MUTATIONS IN EXON 2 OF GATA1 IMPAIR CHROMATIN OCCUPANCY AT ERYTHROID GENES

Research presented in this chapter was carried out in collaboration with the laboratory of Stella Chou. Complementary experiments performed in human cells (discussed in Chapter 5) were executed by Daniel VanDorn and Philip Arca, research technicians in the Chou lab, with bioinformatic expertise provided by Marta Byrska-Bishop, a graduate student in the Ross Hardison laboratory.

This work is included in its entirety in the following manuscript, currently in preparation. Daniel VanDorn, Marta Byrska-Bishop, Amy E. Campbell, Philip R. Arca, Yu Yao, Gerd A. Blobel, Paul Gadue, Fernando Costas, Richard L. Nemiroff, Deborah L. French, Ross C. Hardison, Mitchell J. Weiss, Stella T. Chou. Patient-derived induced pluripotent stem cells reveal distinct hematopoietic defects conferred by truncated GATA1.

Introduction

GATA1 is an essential transcription factor for megakaryocyte and erythroid cell development, where it acts to regulate lineage-specific gene expression (Fujiwara et al., 1996; Shivdasani et al., 1997). In normal hematopoiesis, GATA1 is alternatively spliced to produce two isoforms: a predominant full-length protein that includes exon 2, and a truncated isoform called GATA1s that lacks this exon and therefore the first 83 amino acids (Calligaris et al., 1995) (Figure 1.3). Germline splice site mutations in the GATA1 gene that result in almost exclusive expression of GATA1s have been identified in patients with congenital dyserythropoietic anemia, in which affected individuals exhibit moderate to severe anemia and neutropenia, but no abnormalities in platelet number (Hollanda et al., 2006). More recently, whole exome sequencing of Diamond Blackfan anemia (DBA) patients revealed a similar inherited splice site mutation in GATA1 (Sankaran et al., 2012). In contrast, individuals with trisomy 21 acquire GATA1s mutations in the setting of two clonal hematopoietic disorders, TMD and AMKL (Alford et al., 2011; Wechsler et al., 2002). Notably, euploid patients with anemia due to inherited GATA1s mutations are not predisposed to leukemia.

The N-terminus of GATA1, which is absent in GATA1s, houses a transcriptional activation domain identified by virtue of its ability to turn on reporter gene expression in non-hematopoietic cells (Calligaris et al., 1995; Martin and Orkin, 1990). However, the NAD seems to be dispensable for GATA1 function in several models of hematopoiesis (Kuhl et al., 2005; Weiss et al., 1997). Additionally, and in contrast to the severe anemia and early death seen in GATA1-null mice (Fujiwara et al., 1996), mice genetically engineered to express only GATA1s exhibit anemia and abnormal megakaryopoiesis during fetal development, but have normal adult hematopoiesis (Li et al., 2005). Thus, the function of the GATA1 NAD appears to be context-dependent according to cell type, species, and ontogeny.

In order to better understand the role of the GATA1 NAD in erythromegakaryocytic development and disease, we utilized the G1ME cell gene complementation system (see Chapter

1). Using this system, we find that loss of the GATA1 N-terminus markedly impairs erythroid development and redirects progenitors to a megakaryocytic fate. This lineage-specific defect appears to be due to the fact that GATA1s binding is reduced at erythroid target genes but normal at megakaryocyte genes, an insight that provides a novel mechanism to explain human anemias associated with GATA1s mutations.

Results

The N-terminus of GATA1 is required for normal erythroid maturation

To determine the capacity of N-terminally truncated GATA1 to promote erythroid and megakaryocyte maturation, we studied the function of GATAs in the hematopoietic G1ME cell line. G1ME cells were derived from in vitro differentiation of murine GATA1-null embryonic stem cells and self renew in culture as TPO-dependent, undifferentiated blasts (Stachura et al., 2006). When cultured in the presence of both TPO and EPO, cells reconstituted with GATA1 undergo terminal erythroid and megakaryocyte maturation (Chou et al., 2009; Stachura et al., 2006). We introduced wild-type or mutant GATA1 into G1ME cells and assessed differentiation by morphology and flow cytometry for lineage-specific cell surface markers. Four days after transduction, cells reconstituted with wild-type GATA1 expressed the erythroid and megakaryocyte surface markers Ter119 and Gp1b at a ratio of approximately 3 to 1 (Figure 4.1). These cells had also acquired morphological characteristics of mature erythrocytes, including condensed nuclei and a lightly staining cytoplasm, and mature megakaryocytes, including increased cell size and multilobular nuclei (Figure 4.1). In contrast, G1ME cells expressing GATA1s showed a marked deficiency in erythroid differentiation but robust megakaryocyte maturation (Figure 4.1). They expressed Ter119 and Gp1b at a ratio of about 0.6 to 1. In addition to a reduction in the number of Ter119-positive cells, the intensity of Ter119 expression was reduced. Furthermore, the erythroid cells that were present had large nuclei and a darkly staining

cytoplasm, indicating that they were less mature than their wild-type counterparts. GATA1 and GATA1s-expressing megakaryocytes were indistinguishable.

The effects of GATA1s on erythromegakaryocytic differentiation are consistent with the phenotypes of patients harboring germline GATA1s mutations. These findings validate the use of the G1ME gene complementation system as a faithful and robust model with which to study disease-causing GATA1 mutations.

Differential occupancy of GATA1s at erythroid and megakaryocyte target genes

We next wanted to determine the mechanism(s) by which GATA1s expression leads to defective erythropoiesis. Although the NAD is not required for DNA binding *in vitro* (Martin and Orkin, 1990), several other germline GATA1 mutations also not predicted by *in vitro* or structural studies to impair DNA binding have been shown to affect the ability of GATA1 to bind to target sites *in vivo* (Campbell et al., 2013; Chlon et al., 2012; Letting et al., 2004). Therefore, we hypothesized that loss of the NAD might impair GATA1 chromatin occupancy selectively at erythroid genes. To test this, we carried out ChIP on G1ME cells 48 hours after transduction. At this time point, maturation differences between GATA1 and GATA1s-expressing cells were not yet present (Figure 4.1) and both proteins were expressed equivalently (Figure 4.2). Anti-HA ChIP revealed that GATA1 and GATA1s occupied megakaryocyte genes equivalently, but that GATA1s showed a >40% reduction in binding to erythroid genes (Figure 4.3). Gene expression changes as determined by RT-qPCR analysis were consistent with these occupancy profiles, with erythroid target genes displaying reduced transcriptional activation in GATA1s-expressing cells while megakaryocyte target genes were induced at least as robustly as by wild-type GATA1 (Figure 4.4).

Together, these data suggest that deletion of the first 83 amino acids of GATA1 can diminish GATA1 chromatin occupancy specifically at erythroid target genes, leading to impaired

erythropoiesis. This suggests a new mechanism to help explain the anemic phenotypes found in patients harboring GATA1s mutations.

Figure 4.1. GATA1s impairs erythroid differentiation.

The left panels depict surface expression of the terminal megakaryocyte marker Gp1b and the terminal erythroid marker Ter119 as assessed by flow cytometry in G1ME cells two and four days after infection with GATA1 or GATA1s. Percentages refer to the fraction of GFP-positive cells expressing Gp1b or Ter119. Greater than 90% of transduced cells were GFP-positive. The right panels show MGG staining of G1ME cells expressing GATA1 or GATA1s four days after transduction. Scale bars, 20 μm .

Figure 4.2. Expression of GATA1 and GATA1s in G1ME cells.

Anti-HA and, as a loading control anti-β-actin western blots of extracts from G1ME cells expressing GATA1 or GATA1s 48 hours after transduction. Bands were quantified by densitometry and relative GATA1 levels are noted below the panel. ND = not determined.

Figure 4.3. GATA1s chromatin occupancy at erythroid and megakaryocyte genes.

Anti-HA or anti-IgG (control) ChIP in G1ME cells expressing GATA1 or GATA1s 48 hours after transduction using primers recognizing erythroid- or megakaryocyte-specific target genes. Upper panels depict binding at individual gene targets. Lower panel shows the average binding across all examined erythroid or megakaryocyte (Meg) genes. The occupancy of wild-type GATA1 has been set to 1 in all cases. ** $P < 0.005$. Error bars denote SEM or SD ($n = 4$ for upper panels and $n = 10$ for lower panel).

Figure 4.4. Effect of GATA1s on erythroid and megakaryocyte gene expression.

Expression of GATA1-activated erythroid (top row) and megakaryocyte (bottom row) genes 48 hours after transduction with GATA1 or GATA1s as determined by RT-qPCR, normalized to β -actin, and plotted as fold change from uninfected samples. * P < 0.05. Error bars denote SEM (n = 4).

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS

Myriad human diseases are caused by defects in the regulatory factors that control gene expression. Yet our understanding of the mechanistic consequences of these defects is far from complete. In this work, we focus our attention on the tissue-specific transcription factor GATA1. Mutations in GATA1 cause a spectrum of hematologic disorders, but the mechanisms by which these mutations alter GATA1 function and lead to disease are unclear.

In the first part of this work, we show that missense mutations in the GATA1 NF impair association with the essential cofactors FOG1 and the TAL1 complex. FOG1-disrupting mutations result in greatly reduced transcriptional activation and repression, which severely impairs erythroid and megakaryocyte maturation and leads to pronounced clinical phenotypes. In contrast, mutations that disrupt TAL1 complex binding only moderately inhibit transcriptional activation, resulting in subtle defects to erythroid and megakaryocyte development that cause relatively mild disease presentations. Remarkably, different substitutions at the same residue can selectively inhibit TAL1 complex or FOG1 binding, producing distinct cellular and clinical phenotypes. Perhaps most surprisingly, we reveal that mutations in the DNA-binding surface of the NF do not affect GATA1 occupancy at gene targets *in vivo*, and therefore cause disease through alternate mechanisms (Figure 5.1). (For a summary of all results regarding GATA1 NF mutations, see Table 3.6.)

In the second part of this work, we extend our studies in an effort to understand how splice site mutations in exon 2 of GATA1 cause human disease. Splice site defects lead to the production of an amino-truncated protein called GATA1s, which we demonstrate displays reduced binding at erythroid genes while occupying megakaryocyte-specific genes normally. Reduced erythroid gene occupancy impairs normal red blood cell development and suggests a mechanism for the anemia found in patients with these mutations (Figure 5.2).

In the following sections, these conclusions will be discussed in greater depth and follow-up experiments will be suggested to expand the work in new directions.

A majority of GATA1 NF mutations disrupt FOG1 binding

Four out of seven GATA1 NF mutations disrupt binding to the cofactor FOG1. Our studies establish a strong correlation between binding to full-length FOG1 in vitro, recruitment of FOG1 to target sites in vivo, facilitating GATA1 chromatin occupancy at FOG1-dependent sites, the ability to trigger G1E and G1ME differentiation, and disease severity in patients. This solidifies the idea that reduced FOG1 binding underlies one major class of disease-causing GATA1 mutations. Future efforts could address ways to specifically rescue the effects of FOG1-disrupting mutations, and use this knowledge to create targeted therapies for patients with this type of defect.

Early work from the Orkin laboratory used a split two-hybrid system in yeast to first select for engineered mutations in GATA1 that cripple association with FOG1, and then to find mutations in FOG1 that restore the interaction (Crispino et al., 1999). The GATA1 mutant V205G (an alteration very similar to disease-causing V205M) led to impaired FOG1 binding and erythroid maturation that were rescued by expression of compensatory mutant FOG1 S706R. One could determine if FOG1 S706R restores the differentiation, binding, occupancy, and transcriptional defects caused by GATA1 V205M. If so, does it also rescue G208S, G208R, and D218Y defects? We predict that FOG1 S706R will rescue all of these GATA1 mutants because S706R creates a *de novo* GATA1-FOG1 interaction site rather than specifically binding V205G (Crispino et al., 1999). However, if these GATA1 mutants require unique compensatory mutations in FOG1, a split two-hybrid approach in the model of Crispino et al. could be utilized to find them. In the end, appropriate FOG1 mutants could be introduced into a patient's own HSCs under erythroid- and megakaryocyte-specific promoters that drive expression in the correct cells. Alternatively, genome-editing technologies could be used to directly correct the disease-causing mutations in

the endogenous GATA1 locus. These approaches would eliminate the need for matched donors and reduce complications such as infection and rejection that commonly occur with bone marrow transplant (for now the only curative option for GATA1 mutant patients). Even a moderate restoration of the GATA1-FOG1 interaction could have a marked effect on phenotype, as demonstrated by our comparisons of G208 mutations.

If DNA-binding residues of the GATA1 NF aren't binding DNA, what are they doing?

Given strong in vitro evidence that mutations on the GATA1 NF DNA-interacting surface (R216Q and R216W) impair binding to palindromic and GATC motifs (this study and (Yu et al., 2002)), it was surprising that they do not affect GATA1 target site occupancy in vivo. Similar discrepancies in the context of a distinct GATA1 mutant have been noted (Lamonica et al., 2006). This apparent discordance highlights the importance of validating in vitro studies with in vivo experiments. It remains possible that subtle deficiencies in DNA binding, should they exist, are missed by ChIP-qPCR or are revealed only at sites not included in our analysis. Genome-wide localization of the R216Q and R216W mutants should address these points in future studies. It is also possible that cofactor complexes surrounding GATA1 help stabilize it within chromatin, thereby compensating for the loss of direct DNA binding by the NF, or that the NF simply does not contribute to DNA binding in vivo as much as originally assumed. Importantly, our work provides an alternative explanation for R216Q, which does not impair DNA binding as predicted, but rather inhibits TAL1 complex recruitment to GATA1-occupied genes (see also below).

On the other hand, the mechanism by which R216W affects GATA1 function remains unclear, as we failed to observe any measurable impact on DNA, FOG1, or TAL1 complex binding. There was also no change to the occupancy profiles of several other GATA1 cofactors (CBP/p300 and Cdk9, data not shown). Notably, this mutation does result in a subtle erythroid phenotype in cells – slightly but significantly reducing transcriptional activation and repression.

Therefore, our current working hypothesis is that residue R216 makes contact with an as yet untested or undescribed GATA1 cofactor, and that the R216W mutation selectively disrupts this interaction (Figure 5.1). An untested candidate is EKLF. The lack of a ChIP-grade antibody recognizing endogenous EKLF prevented us from determining if R216W affects EKLF chromatin occupancy. Future studies to address this question could be carried out with exogenously expressed, tagged EKLF. If this yields no definitive answer, we propose an unbiased IP-mass spectrometry screen to identify unknown GATA1-interacting partners whose binding is disrupted by R216W. It is also possible that the R216W mutation creates a binding site for a protein that does not normally associate with GATA1. IP-mass spectrometry could be used address this hypothesis as well.

Our existing data does not explain how the R216W mutation might lead to a presentation of CEP. Gene expression analyses in erythroid cells provided no immediate clue other than a ~50% decrease in *Uros* expression. However, because CEP in humans is associated with much less than 50% UROS activity, this is unlikely to account for the observed patient phenotype. Although it is possible that a lack of any major identifiable deficiencies in R216W-expressing cells reflects a limitation of our assays, it is important to note that this mutation was identified in a single patient with additional genetic abnormalities and co-morbidities, leaving open the possibility that the effects of R216W are influenced by genetic modifiers.

Setting aside patient-derived mutations, more generally there is still an open question as to whether or not any of the structurally defined GATA1 NF DNA-binding residues are important for chromatin occupancy in vivo. Can changes to these residues impair occupancy at palindromic and GATC motifs? To determine this, we could engineer mutations in known DNA-interacting residues such as P213, L214, N226, or Y231, express these altered proteins in G1E cells, and carry out GATA1 ChIP at the same sites interrogated in this document. If occupancy (compared to wild-type GATA1) is decreased at palindromic sites but normal at single motifs, this would suggest the NF might indeed play a role in stabilizing DNA binding in vivo. Unpublished results

describing an anemic and thrombocytopenic “knock-in” mouse harboring a P213G mutation indicate this may be the case (Yu et al., 2002). Of course, ruling out the possibility that any observed effects are indirectly due to reduced cofactor binding is critical.

A new role for the TAL1 complex in GATA1-mediated disease

An unexpected finding from our work is that R216Q and D218G disrupt the interaction between GATA1 and the TAL1 cofactor complex. We therefore implicate the TAL1 complex in non-malignant hematopoietic disease for the first time. This result was not predicted by prior in vitro or structural studies, and is surprising given that R216Q falls on the DNA-binding surface and D218Y disrupts FOG1 association. However, a shared molecular defect is consistent with the similar phenotypes caused by R216Q and D218G mutations. Further, a qualitative difference in TAL1 complex binding between R216Q and R216W, and between D218G and D218Y, explains each pair's divergent patient presentations.

Notably, GATA1 mutations that affect TAL1 complex binding neither measurably diminish nor increase FOG1 recruitment and vice versa, arguing that these proteins do not mutually stabilize or interfere with each other as has been suggested by co-IP experiments (Rodriguez et al., 2005) and consistent with in vitro data showing simultaneous binding of FOG1 and LMO2 to the GATA1 NF (Wilkinson-White et al., 2011).

Relatively few genes are misregulated by the R216Q or D218G mutations, in agreement with prior studies demonstrating limited alterations upon perturbation of the TAL1 complex (Fujiwara et al., 2010; Tripic et al., 2009) and in contrast to FOG1 which changes the expression of substantial numbers of genes. These modest transcriptional effects and the relatively mild phenotype they cause are likely because the TAL1 complex is mostly dispensable for GATA1 chromatin occupancy and gene repression. An outstanding question remains: do these misregulated targets reveal a specific cellular process uniquely affected by disruption of the GATA1-TAL1 complex?

Based on previously published data, one might predict that signatures for erythroid membrane organization or heme biosynthesis would be uncovered (Kassouf et al., 2010). However, preliminary functional annotation and gene set enrichment analyses failed to identify any gene ontology class or pathway uniquely enriched in R216Q or D218G cells relative to other mutant lines. Future studies might expand these bioinformatic efforts in an attempt to define a class of genes whose expression is specifically reliant on the TAL1 complex.

How do different mutations at the same residue selectively disrupt cofactor binding?

A remarkable finding of our study was that distinct substitutions at a single residue can lead to not only quantitative changes (G208S vs. G208R) but also qualitative differences (D218G vs. D218Y and R216Q vs. R216W) in cofactor binding. We speculate that each substitution in R216 or D218 elicits unique phenotypes as a result of different physiochemical properties. Although both R216 mutations remove the positively charged side chain, glutamine substitution trims down the side chain whereas tryptophan substitution introduces a bulky residue that has the potential to form stabilizing cation- π interactions. Thus, the smaller R216Q likely results in minor rearrangements that shift the GATA1-binding surface away from LMO2, whereas the bulkier R216W is able to maintain the surface. Similarly, although both D218 mutations eliminate key hydrogen bonds that likely cause focal changes to the NF structure, conversion to glycine introduces backbone flexibility that could disrupt the LMO2 binding site, whereas mutation to tyrosine introduces a bulky side chain that could preserve the LMO2-binding surface but would necessitate some repacking of the surrounding side chains, disrupting the nearby FOG1-binding surface (Figure 5.3).

A direct test of these predictions would necessitate solving the structure of the GATA1-FOG1-LMO2 interaction, a feat as yet unaccomplished by anyone. This structure could then be compared to those obtained with mutant forms of GATA1 to map the exact changes that occur

between the three proteins in the presence of missense mutations. An alternative approach might be to design GATA1 proteins harboring R216 and D218 mutations that are similar in form to the disease-causing ones, express them in G1E and G1ME cells, and assess maturation, gene expression, chromatin occupancy, and cofactor binding for phenocopy effects. If the structural assumptions made above were true, one would predict that D218A would mimic D218G, D218F would copy D218Y, and R216N would imitate R216Q.

Aberrant GATA1s expression causes erythroid-specific defects

Our work demonstrates that loss of the N-terminus reduces GATA1 occupancy specifically at erythroid target genes, leading to markedly impaired red blood cell development. Future studies will expand these investigations genome-wide and into other cell systems (discussed below).

Contrary to our findings, the NAD has been described as dispensable for GATA1 function in erythrocytes (Weiss et al., 1997). This discrepancy could arise from the particular cell systems used and/or the specific deletions made. The earlier study was carried out in an erythroid-committed cell line, whereas ours was done in bipotential erythromegakaryocytic G1ME cells. This suggests that the NAD plays a critical role during MEP lineage choice to promote the erythroid fate, but may be less important in terminal erythroid maturation. We can test this by expressing GATA1 and GATA1s in G1E cells, and examining cellular maturation, gene expression, and chromatin occupancy profiles. Additionally, the earlier study removed only the first 63 amino acids of the GATA1 N-terminus, compared to our human-disease derived deletion of 83 residues. This suggests that an essential function is housed within these 20 remaining amino acids. In particular, we suspect they might comprise an important protein-protein interaction site. To determine the role of these 20 residues in erythroid differentiation, we can repeat our maturation, gene expression, and chromatin occupancy experiments comparing GATA1, GATA1s(Δ 1-83), and GATA1(Δ 1-63) side-by-side.

The exact mechanisms by which deletion of the NAD reduces GATA1 chromatin occupancy are unclear. We do not expect this domain to have direct effects on DNA binding. Therefore, our working model is that the NAD makes contact with an erythroid-specific cofactor that is important for stabilizing GATA1 binding at target sites and activating transcription (Figure 5.2). When the NAD is missing, this cofactor cannot bind and GATA1 occupancy and gene activation are reduced. Megakaryocytes are not sensitive to loss of the NAD because they do not express this cofactor or use the NAD to facilitate DNA binding. Protein interactions mediated by the NAD remain relatively unexplored. RB is known to bind (Kadri et al., 2009) and has been shown to facilitate an open chromatin state at major histocompatibility complex genes (Osborne et al., 1997). We plan to ChIP for RB in GATA1- and GATA1s-expressing G1MEs at erythroid and megakaryocyte genes to determine if it might play a role in facilitating occupancy at the former. To identify other potential NAD-associated factors, in the future we plan to carry out an unbiased IP-mass spectrometry screen to identify proteins that fail to associate with GATA1s, but bind GATA1 in erythroid cells. If residues 64-83 are found to be important in the proposed maturation studies, this will further inform our interaction screen.

Can individual gene studies be extrapolated genome-wide?

Almost all of the studies discussed in this document were carried out by examining expression of and factor occupancy at several dozen GATA1-regulated genes. We fully expect the conclusions made based on these individual targets to hold up when global analyses are carried out.

For GATA1 NF mutants, expression profiling on a wide scale by microarray has been completed and supports results from RT-qPCR experiments (see Chapter 3). Future GATA1, FOG1, LMO2, and TAL1 ChIP-sequencing will establish the models proposed here on a whole genome level. We do keep in mind that in looking genome-wide we might also discover something novel, as was the case for a recently published study of GATA1 occupancy in

megakaryocytes (Chlon et al., 2012). Only once these high-throughput sequencing experiments are done will we know for sure.

As a continuation of our GATA1s studies, we are examining the transcriptomes of GATA1- and GATA1s-expressing cells harvested two days after transduction by microarray. Our preliminary results support the conclusions made after inspecting only a handful of GATA1 targets; namely, that erythroid genes display reduced transcriptional activation in GATA1s-expressing cells while megakaryocyte genes are induced more robustly (Figure 5.4). We are also in the process of examining GATA1 and GATA1s occupancy genome-wide by ChIP-sequencing. Once we have this data, we will carry out a differential binding analysis using the DiffBind package in R to determine if there is statistically significant difference in binding between GATA1 and GATA1s at erythroid genes but not megakaryocyte genes at the whole genome level, consistent with what we observe by ChIP-qPCR.

Systems for studying disease-causing mutations

Previous studies of GATA1 NF mutations relied mostly on in vitro assays (Freson et al., 2001; 2002; 2003; Mehaffey et al., 2001; Nichols et al., 2000; Yu et al., 2002), with only two mutants (V205M and R216Q) examined in any cell-based system (Letting et al., 2004; Nichols et al., 2000; Yu et al., 2002). In contrast, prior studies of GATA1s, though limited, were predominantly carried out in vivo (Kuhl et al., 2005; Li et al., 2005; Martin and Orkin, 1990; Weiss et al., 1997). Notably, all of the work to date, including research described in this document, has been done in murine systems. As our efforts here demonstrate, it is vitally important to pair any in vitro studies of disease-causing mutations with cellular and/or animal models. But which systems are best? In the case of GATA1, the G1E and G1ME gene complementation assays mimic patient phenotypes quite faithfully. These systems afford a homogeneous cellular and genetic background that allows for the detection of subtle defects that may be masked in studies of heterogeneous patient

samples; and it is also quite easy to collect large quantities of material for biochemical experiments and whole genome analyses. Using these two cell lines, we were able to uncover previously unappreciated mechanisms for GATA1 mutations that were not predicted from in vitro studies or structural considerations. Despite their demonstrated usefulness, the G1E and G1ME systems have some limitations as reflected in subtle phenotypic discrepancies between cells and patients. For example, the G208S mutation causes mild dyserythropoiesis without anemia in patients but in G1E cells produces more pronounced defects (see Chapter 3 and (Mehaffey et al., 2001)). Furthermore, the mild thalassemia seen in some R216Q patients is not recapitulated in G1E cells (see Chapter 3 and (Yu et al., 2002)). These variations might be rooted in species differences or the fact that these systems do not recapitulate all stages of erythromegakaryocytic maturation. Future experiments confirming the results from G1E and G1ME cells in whole animal models and/or human systems would increase confidence in our proposed models. Below are some preliminary results from human systems that validate our findings from murine cell lines.

We first attempted to corroborate our results using primary hematopoietic material obtained from affected human individuals. This was a challenge, as diseases caused by GATA1 mutations are rare and therefore so are patient samples. Compounding this was the reality that several patients had passed away or undergone curative bone marrow transplants. Furthermore, we were unable to request collection of required material from individuals with mild phenotypes as it was medically unnecessary. Finally, several samples that were available were of inadequate quality to be of use in cell culture experiments. Even if we had successfully obtained primary tissue, it would have been in quantities insufficient to carry out key biochemical and ChIP experiments needed to confirm molecular mechanisms suggested by the G1E and G1ME experiments.

In lieu of primary hematopoietic material, we sought to generate induced pluripotent stem cells (iPSCs) from patient fibroblasts or peripheral blood that could be differentiated in culture for further study. The laboratory of Dr. Stella Chou at The Children's Hospital of Philadelphia carried out these iPSC experiments. All iPSCs derived from a patient harboring the V205M mutation

displayed numerous, unique karyotypic abnormalities that were acquired during reprogramming. These abnormal iPSCs were not analyzed further. In contrast, we were able to generate karyotypically normal iPSCs from peripheral blood-derived mononuclear cells from one patient with a germline GATA1 exon 2 mutation. To determine the effect of GATA1s on hematopoiesis in this system, we generated embryoid bodies (EBs) in defined media containing sequential combinations of cytokines to induce primitive streak/mesoderm formation followed by hematopoietic specification (Chou et al., 2012) (Figure 5.5A). On days 7-8 of EB differentiation, hematopoietic progenitors co-expressing CD43⁺41⁺235⁺ were present in EB cultures (Figure 5.5B), and by day 12 of differentiation monolineage hematopoietic cells released from EBs were erythroid (CD235⁺41⁻), myeloid (CD45⁺18⁺), or megakaryocytic (CD41⁺42⁺) (Figure 5.5C). GATA1s iPSCs produced multipotent progenitor populations (CD43⁺41⁺235⁺) at similar levels compared to lines with wild-type GATA1 (Figure 5.5B,D), but GATA1s progenitors failed to differentiate into CD235⁺41⁻ erythroblasts despite retained capacity for myeloid and megakaryocytic differentiation (Figure 5.5C,E). Morphologic examination of suspension cells of day 20 EB cultures showed numerous late stage erythroblasts derived from GATA1 iPSCs that were lacking in cultures derived from GATA1s iPSCs (Figure 5.5F).

In summary, we obtained very similar results regarding the effect of GATA1s on erythroid maturation from both human iPSCs and murine G1ME cells. Whether comparable gene expression changes and occupancy profiles are present in the human cells is as yet undetermined. While the former question can be addressed with microarray or single-cell PCR analyses, we are unable to generate sufficient quantities of material for the ChIP experiments that would address the latter. With current technologies, this is an inherent limitation of the iPSC system. Another drawback to this model is that hematopoietic developmental potential varies between iPSC clones, likely due to genetic differences in the subjects from which the iPSCs are derived (Choi et al., 2009; Mills et al., 2013). This may render the study of mutations that produce subtle phenotypes (like R216Q, R216W, and D218G) a challenge. In terms of GATA1s, this variation requires that iPSCs be derived from additional affected individuals before strong

conclusions are drawn. In all cases, it would be most rigorous to use genome editing to generate isogenic iPSC lines that differ solely at the disease-associated nucleotide (Soldner et al., 2011). This engineering-based approach may present a solution to the problem of unavailable GATA1 NF mutant patient samples, and allow us to create iPSC lines that recapitulate any described GATA1 mutation.

Do other germline GATA1 mutations exist?

The frequency of germline GATA1 mutations is unknown. Since the first description in 2000, less than 15 families have been reported. Are these lesions extremely rare, or are physicians failing to identify them? Until knowledge of GATA1-mediated disorders becomes more widespread in the hematology community, they are liable to be underreported. Further, patients presenting with symptoms (especially mild ones such as those caused by defects in LMO2 binding) but lacking a clear X-linked inheritance pattern are likely to escape correct genetic diagnosis. Therefore, it is probable that many individuals with unexplained hematologic defects are harboring undiscovered GATA1 mutations. Accurately identifying people with GATA1 mutations allows clinicians to better estimate their disease prognosis and prevents unnecessary treatments such as immune therapy, splenectomy and steroids, to which GATA1 mutant patients are usually unresponsive. New knowledge and improving technologies should allow future hematologists to tailor patient care to mutation status.

Moving forward, targeted sequencing of the GATA1 gene should be routinely incorporated into the workups of individuals with anemia and thrombocytopenia of unknown origin. Beyond this candidate gene approach, the increasing clinical application of whole-exome and whole-genome sequencing technologies promises to uncover GATA1 mutations medical practitioners did not even know they were looking for. The discovery of new GATA1 mutations may lead to the identification of novel GATA1 cofactors or interaction surfaces, telling us as much about basic

biology as disease. Of note, all currently identified defects in GATA1 cause discrete changes to protein function, leaving relatively strong residual activities. We hypothesize that this will be true for all future discoveries as well. More drastic alterations, such as mutations in the primary DNA-binding CF domain, would very likely not be compatible with life.

Despite hurdles to diagnosis, astute clinicians have recently used targeted sequencing to find unique GATA1 mutations. David Anstee's group in the United Kingdom recently described a missense mutation in the termination codon (Singleton et al., 2013), the first ever noted alteration in the C-terminal portion of GATA1. We hypothesize that this mutation extends the GATA1 open reading frame, creating an elongated protein with altered cofactor binding potential. Additionally, just last month Kathleen Freson and colleagues in Belgium reported a third missense mutation in codon 218, one that changes the normal aspartic acid to an asparagine (Hermans et al., 2013). Previously described mutations at this residue disrupt FOG1 binding (D218Y) or LMO2 binding (D218G), and lead to severe or moderate phenotypes. Based on the mild thrombocytopenia described by Hermans et al., we would hypothesize that D218N mimics D218G and selectively disrupts LMO2 binding. Future experiments will utilize the G1E and G1ME gene complementation systems to address these two hypotheses.

Whole-exome sequencing is also already demonstrating that additional GATA1 mutations exist. Vijay Sankaran and Hanna Gazda at Harvard Medical School sequenced DNA from ~60 DBA patients without known pathogenic lesions and uncovered two additional germline exon 2 splice site changes that lead to production of amino-truncated GATA1 (Sankaran et al., 2012). The work on GATA1s presented in this document will help inform the molecular etiology of anemias caused by these splice site mutations.

What can we learn about mutations in other GATA family members?

In addition to GATA1, there are five other GATA proteins. GATA2 and GATA3 are expressed predominantly within the hematopoietic system, while GATA4, GATA5, and GATA6 are expressed outside of the blood in mesoderm and endoderm derived tissues such as the heart, intestine, and lung. The zinc finger regions are highly homologous between GATA factors, though conservation is low throughout the rest of the protein (Orkin, 1992). Like GATA1, GATA2, GATA3, GATA4, GATA5 and GATA6 associate with FOG1 (or the closely related FOG2) (Cantor and Orkin, 2005). Additionally, GATA2 and GATA3 have been shown to bind to the TAL1 complex (Ono et al., 1998; Osada et al., 1995). Because the NF itself and its cofactor interactions are conserved, what we have learned about GATA1 protein interactions through our studies may be transferable to other GATA factors and their disease-causing mutations. Future investigations could explore the effects of NF mutations in GATA2 (Dickinson et al., 2011; Greif et al., 2012; Hsu et al., 2011; Ostergaard et al., 2011; Pasquet et al., 2013), GATA3 (Gaynor et al., 2009; Nesbit et al., 2004), GATA4 (Lourenço et al., 2011; Reamon-Buettner and Borlak, 2005), or GATA5 (Gu et al., 2012; Wei et al., 2013) on DNA and cofactor binding to better link molecular mechanism to disease phenotype.

In summary, this work sheds new light on the pathophysiology and mechanisms underlying congenital hematologic diseases caused by mutations in the essential, tissue-specific transcription factor GATA1. This deeper understanding should improve classification, refine clinical management of affected individuals, and allow prediction of disease progression based on molecular defect. More broadly, we provide a paradigm for better understanding disease-causing mutations through a combined modality approach. Because mechanisms predicted by in vitro studies may not always reflect events in biologically relevant contexts, in vivo studies such as those reported here serve to not only validate predictions but also provide entirely new and unexpected insights. Our approach could be applied in other contexts where gene mutations lead

to variable clinical presentations, and underlying effects on protein function are incompletely understood.

Figure 5.1. GATA1 NF mutations disrupt cofactor binding: a summary model.

Wild-type GATA1 binds chromatin, activating the transcription of lineage-specific genes (arrow at right) and repressing transcription of genes associated with proliferative states and alternative lineages (bar at left), leading to normal erythroid and megakaryocyte maturation. When mutations (green star) disrupt FOG1 binding (V205M, G208S, G208R, D218Y), chromatin occupancy is reduced at some binding sites, transcriptional activation and repression are greatly impaired (green X), and erythroid and megakaryocyte maturation are severely affected, leading to deep anemia and thrombocytopenia. When mutations disrupt TAL1 complex binding (R216Q, D218G), transcriptional activation is moderately impaired and erythroid and megakaryocyte maturation are

somewhat inhibited which leads to anemia and thrombocytopenia. We hypothesize that the R216W mutation disrupts binding to an unknown cofactor (question mark), mildly reducing transcriptional activation and repression to cause erythroid porphyria and thrombocytopenia. NF, N-terminal zinc finger; CF, C-terminal zinc finger.

Figure 5.2. GATA1s impairs erythropoiesis: a summary model.

Wild-type GATA1 (left panel) associates via the NAD with an unknown erythroid-specific factor important for stable GATA1 chromatin occupancy and erythroid gene activation. In this cellular context, normal maturation of both red blood cells and megakaryocytes occurs. Mutant GATA1s (right panel) lacks the NAD and therefore fails to bind the aforementioned erythroid-specific factor. This leads to reduced GATA1 chromatin occupancy and transcriptional activation at erythroid genes, impairing red blood cell development. NAD, N-terminal activation domain; MK, megakaryocyte.

Figure 5.3. Possible physical causes of differential effects of R216 and D218 mutations.

GATA1 NF (grey; surface representation), FOG1 (cyan; surface representation), LMO2-Ldb1 (orange; surface representation). **(A-C)** R216 mutations. Wild-type R216 shown in red, Q216 shown in purple, and W216 shown in magenta. Whereas the R216W mutation is similar in size to the wild-type side chain, and may be stabilized by various new interactions as indicated (dashed lines), the R216Q mutation is smaller leading to possible collapse of the LMO2-binding face (black arrow) due to repacking of side chains. **(D-F)** D218 mutations. Wild-type D218 (red), G218 (purple), and Y218 (magenta). D218G truncates the side chain and introduces flexibility into the backbone, leading to possible collapse of the LMO2-binding face (black arrow) due to repacking of the side chains. The larger D218Y mutation could maintain the position of the LMO2-binding face, but require repacking of the small hydrophobic core of the NF (black arrow) leading to disruption of the FOG1-binding face. Models are based on the coordinates of PDB accession code 2L6Z. NF, N-terminal zinc finger; WT, wild-type.

Figure 5.4. Gene set enrichment analysis of G1ME microarray data.

Gene set enrichment analysis was performed with erythroid (Ery4 and Ery5) and megakaryocyte (Mega2) gene sets obtained from the studies of (Novershtern et al., 2011). Left panel: genes that were ≥ 2 -fold upregulated in GATA1s-expressing cells as compared to wild-type GATA1-expressing cells. Right panel: genes that were ≥ 2 -fold downregulated in GATA1s-expressing cells as compared to GATA1-expressing cells. Genes were ranked according to their correlation with either the erythroid, “Ery”, phenotype (red) or megakaryocyte, “Meg”, phenotype (blue). The left panel shows that genes upregulated in GATA1s-expressing cells were significantly enriched for megakaryocyte genes. The right panel shows that genes downregulated in GATA1s-expressing cells were significantly enriched in erythroid genes. ES, enrichment score.

Figure 5.5. Absent erythropoiesis from iPSCs harboring GATA1s mutations.

(A) Hematopoietic differentiation protocol. (B) Flow cytometry analysis showing $CD34^{+/-}43^{+}235^{+}41^{+}$ progenitors within total embryoid body (EB) cultures on day 7 of hematopoietic differentiation and (C) suspension cells released from EBs on day 12 showing mature hematopoietic lineages: erythroid ($CD41^{-}235^{+}$), megakaryocytic (Meg, $CD41^{+}42^{+}$), and myeloid ($CD45^{+}18^{+}$). (D) Frequency of $CD43^{+}41^{+}235^{+}$ progenitor cells in EB cultures on day 7-8 of hematopoietic differentiation. (E) Summary of distribution of lineage-committed cells in EB suspension cultures at day 12-14 of differentiation. (F) Morphology of hematopoietic cells in day 20 differentiation cultures.

REFERENCES

- Alford, K.A., Reinhardt, K., Garnett, C., Norton, A., Böhmer, K., Neuhoﬀ, von, C., Kolenova, A., Marchi, E., Klusmann, J.-H., Roberts, I., et al. (2011). Analysis of GATA1 mutations in Down syndrome transient myeloproliferative disorder and myeloid leukemia. *Blood* 118, 2222–2238.
- Athanasiou, M., Mavrothalassitis, G., Sun-Hoﬀman, L., and Blair, D.G. (2000). FLI-1 is a suppressor of erythroid diﬀerentiation in human hematopoietic cells. *Leukemia* 14, 439–445.
- Ayed, A., Mulder, F.A., Yi, G.S., Lu, Y., Kay, L.E., and Arrowsmith, C.H. (2001). Latent and active p53 are identical in conformation. *Nat Struct Biol* 8, 756–760.
- Balduini, C.L., Pecci, A., Loffredo, G., Izzo, P., Noris, P., Grosso, M., Bergamaschi, G., Rosti, V., Magrini, U., Ceresa, I.F., et al. (2004). Eﬀects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. *Thromb Haemost* 91, 129–140.
- Balwani, M., and Desnick, R.J. (2012). The porphyrias: advances in diagnosis and treatment. *Blood* 120, 4496–4504.
- Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.-P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., et al. (2012). Exome sequencing identiﬁes recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 44, 685–689.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289–300.
- Berman, B.P., Nibu, Y., Pfeiﬀer, B.D., Tomancak, P., Celniker, S.E., Levine, M., Rubin, G.M., and Eisen, M.B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc Natl Acad Sci USA* 99, 757–762.
- Blobel, G.A., Nakajima, T., Eckner, R., Montminy, M., and Orkin, S.H. (1998). CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid diﬀerentiation. *Proc Natl Acad Sci USA* 95, 2061–2066.
- Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998). Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* 396, 594–598.
- Bradsher, J.N., Tan, S., McLaury, H.J., Conaway, J.W., and Conaway, R.C. (1993). RNA polymerase II transcription factor SIII. II. Functional properties and role in RNA chain elongation. *J Biol Chem* 268, 25594–25603.
- Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I., and Santoro, C. (1995). Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. *Proc Natl Acad Sci USA* 92, 11598–11602.
- Campbell, A.E., Wilkinson-White, L., Mackay, J.P., Matthews, J.M., and Blobel, G.A. (2013). Analysis of disease-causing GATA1 mutations in murine gene complementation systems. *Blood* 121, 5218–5227.

- Cantor, A.B., and Orkin, S.H. (2005). Coregulation of GATA factors by the Friend of GATA (FOG) family of multitype zinc finger proteins. *Semin Cell Dev Biol* 16, 117–128.
- Cantor, A.B., Iwasaki, H., Arinobu, Y., Moran, T.B., Shigematsu, H., Sullivan, M.R., Akashi, K., and Orkin, S.H. (2008). Antagonism of FOG-1 and GATA factors in fate choice for the mast cell lineage. *J Exp Med* 205, 611–624.
- Chang, A.N., Cantor, A.B., Fujiwara, Y., Lodish, M.B., Droho, S., Crispino, J.D., and Orkin, S.H. (2002). GATA-factor dependence of the multitype zinc-finger protein FOG-1 for its essential role in megakaryopoiesis. *Proc Natl Acad Sci USA* 99, 9237–9242.
- Cheng, Y., Wu, W., Kumar, S.A., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.-B., Zhang, Y., Drautz, D., et al. (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res* 19, 2172–2184.
- Chlon, T.M., Dore, L.C., and Crispino, J.D. (2012). Cofactor-mediated restriction of GATA-1 chromatin occupancy coordinates lineage-specific gene expression. *Molecular Cell* 47, 608–621.
- Choi, K.-D., Yu, J., Smuga-Otto, K., Salvagiotto, G., Rehrauer, W., Vodyanik, M., Thomson, J., and Slukvin, I. (2009). Hematopoietic and endothelial differentiation of human induced pluripotent stem cells. *Stem Cells* 27, 559–567.
- Chou, S.T., Byrsk-Bishop, M., Tober, J.M., Yao, Y., Vandorn, D., Opalinska, J.B., Mills, J.A., Choi, J.K., Speck, N.A., Gadue, P., et al. (2012). Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc Natl Acad Sci USA* 109, 17573–17578.
- Chou, S.T., Khandros, E., Bailey, L.C., Nichols, K.E., Vakoc, C.R., Yao, Y., Huang, Z., Crispino, J.D., Hardison, R.C., Blobel, G.A., et al. (2009). Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate. *Blood* 114, 983–994.
- Ciovacco, W.A., Raskind, W.H., and Kacena, M.A. (2008). Human phenotypes associated with GATA-1 mutations. *Gene* 427, 1–6.
- Clark, A.J., Doyle, K.M., and Humbert, P.O. (2004). Cell-intrinsic requirement for pRb in erythropoiesis. *Blood* 104, 1324–1326.
- Conaway, R.C., and Conaway, J.W. (2009). The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem Sci* 34, 71–77.
- Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* 2, 440–452.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848.
- Crispino, J.D., Lodish, M.B., Mackay, J.P., and Orkin, S.H. (1999). Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Molecular Cell* 3, 219–228.
- Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12–27.

- Dawson, M.A., Prinjha, R.K., Dittman, A., Giotopoulos, G., Bantscheff, M., Chan, W.-I., Robson, S.C., Chung, C.-W., Hopf, C., Savitski, M.M., et al. (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529–533.
- Del Vecchio, G.C., Giordani, L., De Santis, A., and De Mattia, D. (2005). Dyserythropoietic anemia and thrombocytopenia due to a novel mutation in GATA-1. *Acta Haematol* **114**, 113–116.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244.
- Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3.
- Deutsch, V.R., and Tomer, A. (2006). Megakaryocyte development and platelet production. *Br J Haematol* **134**, 453–466.
- Dey, A., Ellenberg, J., Farina, A., Coleman, A.E., Maruyama, T., Sciortino, S., Lippincott-Schwartz, J., and Ozato, K. (2000). A bromodomain protein, MCAP, associates with mitotic chromosomes and affects G2-to-M transition. *Mol Cell Biol* **20**, 6537–6549.
- Dey, A., Chitsaz, F., Abbasi, A., Misteli, T., and Ozato, K. (2003). The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc Natl Acad Sci USA* **100**, 8758–8763.
- Dey, A., Nishiyama, A., Karpova, T., McNally, J., and Ozato, K. (2009). Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. *Mol Biol Cell* **20**, 4899–4909.
- Dickinson, R.E., Griffin, H., Bigley, V., Reynard, L.N., Hussain, R., Haniffa, M., Lakey, J.H., Rahman, T., Wang, X.-N., McGovern, N., et al. (2011). Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* **118**, 2656–2658.
- Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**, 1475–1489.
- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* **489**, 101–108.
- Duan, Z., and Horwitz, M. (2003). Gfi-1 oncoproteins in hematopoiesis. *Hematology* **8**, 339–344.
- Dührsen, U., Kratz, C.P., Flotho, C., Lauenstein, T., Bommer, M., König, E., Brittinger, G., and Heimpel, H. (2011). Long-term outcome of hemizygous and heterozygous carriers of a germline GATA1 (G208R) mutation. *Ann Hematol* **90**, 301–306.
- Dzierzak, E., and Philipsen, S. (2013). Erythropoiesis: Development and Differentiation. *Cold Spring Harbor Perspectives in Medicine* **3**, a011601–a011601.
- Eisbacher, M., Holmes, M.L., Newton, A., Hogg, P.J., Khachigian, L.M., Crossley, M., and Chong, B.H. (2003). Protein-protein interaction between Fli-1 and GATA-1 mediates synergistic

expression of megakaryocyte-specific genes through cooperative DNA binding. *Mol Cell Biol* 23, 3427–3441.

Elagib, K.E., Racke, F.K., Mogass, M., Khetawat, R., Delehanty, L.L., and Goldfarb, A.N. (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* 101, 4333–4341.

Evans, T., and Felsenfeld, G. (1989). The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell* 58, 877–885.

Evans, T., Reitman, M., and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci USA* 85, 5976–5980.

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol* 25, 1215–1227.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. *Nature* 468, 1067–1073.

Forsberg, E.C., Downs, K.M., Christensen, H.M., Im, H., Nuzzi, P.A., and Bresnick, E.H. (2000). Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc Natl Acad Sci USA* 97, 14494–14499.

Fox, A.H., Kowalski, K., King, G.F., Mackay, J.P., and Crossley, M. (1998). Key residues characteristic of GATA N-fingers are recognized by FOG. *J Biol Chem* 273, 33595–33603.

Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *Embo J* 18, 2812–2822.

French, C.A., Ramirez, C.L., Kolmakova, J., Hickman, T.T., Cameron, M.J., Thyne, M.E., Kutok, J.L., Toretsky, J.A., Tadavarthy, A.K., Kees, U.R., et al. (2008). BRD-NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene* 27, 2237–2242.

Freson, K., Devriendt, K., Matthijs, G., Van Hoof, A., De Vos, R., Thys, C., Minner, K., Hoylaerts, M.F., Vermynen, J., and Van Geet, C. (2001). Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood* 98, 85–92.

Freson, K., Matthijs, G., Thys, C., Mariën, P., Hoylaerts, M.F., Vermynen, J., and Van Geet, C. (2002). Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. *Hum Mol Genet* 11, 147–152.

Freson, K., Thys, C., Wittewrongel, C., Vermynen, J., Hoylaerts, M.F., and Van Geet, C. (2003). Molecular cloning and characterization of the GATA1 cofactor human FOG1 and assessment of its binding to GATA1 proteins carrying D218 substitutions. *Hum Genet* 112, 42–49.

Frontelo, P., Manwani, D., Galdass, M., Karsunky, H., Lohmann, F., Gallagher, P.G., and Bieker, J.J. (2007). Novel role for EKLF in megakaryocyte lineage commitment. *Blood* 110, 3871–3880.

- Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192.
- Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* **24**, 787–795.
- Fujiwara, T., Lee, H.-Y., Sanalkumar, R., and Bresnick, E.H. (2010). Building multifunctionality into a complex containing master regulators of hematopoiesis. *Proc Natl Acad Sci USA* **107**, 20429–20434.
- Fujiwara, Y., Browne, C.P., Cunliffe, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci USA* **93**, 12355–12358.
- Gaj, T., Gersbach, C.A., and Barbas, C.F., III (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31**, 397–405.
- Gao, X., Yang, J., Tsang, J.C.H., Ooi, J., Wu, D., and Liu, P. (2013). Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. *Stem Cell Reports* **1**, 183–197.
- Garriga-Canut, M., Agustín-Pavón, C., Herrmann, F., Sánchez, A., Dierssen, M., Fillat, C., and Isalan, M. (2012). Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proc Natl Acad Sci USA* **109**, E3136–E3145.
- Gaynor, K.U., Grigorieva, I.V., Nesbit, M.A., Cranston, T., Gomes, T., Gortner, L., and Thakker, R.V. (2009). A missense GATA3 mutation, Thr272Ile, causes the hypoparathyroidism, deafness, and renal dysplasia syndrome. *J Clin Endocrinol Metab* **94**, 3897–3904.
- Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451.
- Goardon, N., Lambert, J.A., Rodriguez, P., Nissaire, P., Herblot, S., Thibault, P., Dumenil, D., Strouboulis, J., Romeo, P.-H., and Hoang, T. (2006). ETO2 coordinates cellular proliferation and differentiation during erythropoiesis. *Embo J* **25**, 357–366.
- Gregory, G.D., Miccio, A., Bersenev, A., Wang, Y., Hong, W., Zhang, Z., Poncz, M., Tong, W., and Blobel, G.A. (2010). FOG1 requires NuRD to promote hematopoiesis and maintain lineage fidelity within the megakaryocytic-erythroid compartment. *Blood* **115**, 2156–2166.
- Greif, P.A., Dufour, A., Konstandin, N.P., Ksienzyk, B., Zellmeier, E., Tizazu, B., Sturm, J., Benthaus, T., Herold, T., Yaghmaie, M., et al. (2012). GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood* **120**, 395–403.
- Grimes, H.L., Chan, T.O., Zweidler-McKay, P.A., Tong, B., and Tschlis, P.N. (1996). The Gfi-1 proto-oncoprotein contains a novel transcriptional repressor domain, SNAG, and inhibits G1 arrest induced by interleukin-2 withdrawal. *Mol Cell Biol* **16**, 6263–6272.
- Gu, J.-Y., Xu, J.-H., Yu, H., and Yang, Y.-Q. (2012). Novel GATA5 loss-of-function mutations underlie familial atrial fibrillation. *Clinics (Sao Paulo)* **67**, 1393–1399.

Hall, M.A., Curtis, D.J., Metcalf, D., Elefanty, A.G., Sourris, K., Robb, L., Gothert, J.R., Jane, S.M., and Begley, C.G. (2003). The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci USA* 100, 992–997.

Hall, M.A., Slater, N.J., Begley, C.G., Salmon, J.M., Van Stekelenburg, L.J., McCormack, M.P., Jane, S.M., and Curtis, D.J. (2005). Functional but abnormal adult erythropoiesis in the absence of the stem cell leukemia gene. *Mol Cell Biol* 25, 6355–6362.

Hart, A., Melet, F., Grossfeld, P., Chien, K., Jones, C., Tunnacliffe, A., Favier, R., and Bernstein, A. (2000). Fli-1 is required for murine vascular and megakaryocytic development and is hemizyously deleted in patients with thrombocytopenia. *Immunity* 13, 167–177.

Hattangadi, S.M., Wong, P., Zhang, L., Flygare, J., and Lodish, H.F. (2011). From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood* 118, 6258–6268.

Heller, P.G., Glembotsky, A.C., Gandhi, M.J., Cummings, C.L., Pirola, C.J., Marta, R.F., Kornblihtt, L.I., Drachman, J.G., and Molinas, F.C. (2005). Low Mpl receptor expression in a pedigree with familial platelet disorder with predisposition to acute myelogenous leukemia and a novel AML1 mutation. *Blood* 105, 4664–4670.

Hermans, C., De Waele, L., Van Geet, C., and Freson, K. (2013). Novel GATA1 mutation in residue D218 leads to macrothrombocytopenia and clinical bleeding problems. *Platelets* Aug 23. [Epub ahead of print].

Hollanda, L.M., Lima, C.S.P., Cunha, A.F., Albuquerque, D.M., Vassallo, J., Ozelo, M.C., Joazeiro, P.P., Saad, S.T.O., and Costa, F.F. (2006). An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat Genet* 38, 807–812.

Hong, W., Nakazawa, M., Chen, Y.-Y., Kori, R., Vakoc, C.R., Rakowski, C., and Blobel, G.A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *Embo J* 24, 2367–2378.

Hromas, R., Orazi, A., Neiman, R.S., Maki, R., Van Beveran, C., Moore, J., and Klemsz, M. (1993). Hematopoietic lineage- and stage-restricted expression of the ETS oncogene family member PU.1. *Blood* 82, 2998–3004.

Hsu, A.P., Sampaio, E.P., Khan, J., Calvo, K.R., Lemieux, J.E., Patel, S.Y., Frucht, D.M., Vinh, D.C., Auth, R.D., Freeman, A.F., et al. (2011). Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* 118, 2653–2655.

Huang, D.-Y., Kuo, Y.-Y., and Chang, Z.-F. (2005). GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. *Nucleic Acids Res* 33, 5331–5342.

Hung, H.L., Lau, J., Kim, A.Y., Weiss, M.J., and Blobel, G.A. (1999). CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Mol Cell Biol* 19, 3496–3505.

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Ogawa, S., Kurokawa, M., et al. (2004). AML-1 is required for megakaryocytic maturation and

lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat Med* 10, 299–304.

Jacks, T., Fazeli, A., Schmitt, E.M., Bronson, R.T., Goodell, M.A., and Weinberg, R.A. (1992). Effects of an Rb mutation in the mouse. *Nature* 359, 295–300.

Jackson, C.W. (1973). Cholinesterase as a possible marker for early cells of the megakaryocytic series. *Blood* 42, 413–421.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318–356.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular Cell* 19, 523–534.

Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Molecular Cell* 29, 232–242.

Johnson, K.D., Boyer, M.E., Kang, J.-A., Wickrema, A., Cantor, A.B., and Bresnick, E.H. (2007). Friend of GATA-1-independent transcriptional repression: a novel mode of GATA-1 function. *Blood* 109, 5230–5233.

Kadri, Z., Shimizu, R., Ohneda, O., Maouche-Chretien, L., Gisselbrecht, S., Yamamoto, M., Romeo, P.-H., Leboulch, P., Chretien, S., and Goodell, M.A. (2009). Direct binding of pRb/E2F-2 to GATA-1 regulates maturation and terminal cell division during erythropoiesis. *PLoS Biol* 7, e1000123.

Kaneko, H., Kobayashi, E., Yamamoto, M., and Shimizu, R. (2012). N- and C-terminal transactivation domains of GATA1 protein coordinate hematopoietic program. *J Biol Chem* 287, 21439–21449.

Kassouf, M.T., Hughes, J.R., Taylor, S., McGowan, S.J., Soneji, S., Green, A.L., Vyas, P., and Porcher, C. (2010). Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. *Genome Res* 20, 1064–1083.

Kaufmann, R., Straussberg, R., Mandel, H., Fattal-Valevski, A., Ben-Zeev, B., Naamati, A., Shaag, A., Zenvirt, S., Konen, O., Mimouni-Bloch, A., et al. (2010). Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am J Hum Genet* 87, 667–670.

Kawada, H., Ito, T., Pharr, P.N., Spyropoulos, D.D., Watson, D.K., and Ogawa, M. (2001). Defective megakaryopoiesis and abnormal erythroid development in Fli-1 gene-targeted mice. *Int J Hematol* 73, 463–468.

Kihm, A.J., Kong, Y., Hong, W., Russell, J.E., Rouda, S., Adachi, K., Simon, M.C., Blobel, G.A., and Weiss, M.J. (2002). An abundant erythroid protein that stabilizes free alpha-haemoglobin. *Nature* 417, 758–763.

Kim, J., Guermah, M., and Roeder, R.G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* 140, 491–503.

- Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 442, 312–316.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Kratz, C.P., Niemeyer, C.M., Karow, A., Volz-Fleckenstein, M., Schmitt-Gräff, A., and Strahm, B. (2008). Congenital transfusion-dependent anemia and thrombocytopenia with myelodysplasia due to a recurrent GATA1(G208R) germline mutation. *Leukemia* 22, 432–434.
- Kuhl, C., Atzberger, A., Iborra, F., Nieswandt, B., Porcher, C., and Vyas, P. (2005). GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. *Mol Cell Biol* 25, 8592–8606.
- Lamonica, J.M., Deng, W., Kadauke, S., Campbell, A.E., Gamsjaeger, R., Wang, H., Cheng, Y., Billin, A.N., Hardison, R.C., Mackay, J.P., et al. (2011). Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proc Natl Acad Sci USA* 108, E159–E168.
- Lamonica, J.M., Vakoc, C.R., and Blobel, G.A. (2006). Acetylation of GATA-1 is required for chromatin occupancy. *Blood* 108, 3736–3738.
- Lee, E.Y., Chang, C.Y., Hu, N., Wang, Y.C., Lai, C.C., Herrup, K., Lee, W.H., and Bradley, A. (1992). Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. *Nature* 359, 288–294.
- Lee, H.-Y., Johnson, K.D., Fujiwara, T., Boyer, M.E., Kim, S.-I., and Bresnick, E.H. (2009). Controlling hematopoiesis through sumoylation-dependent regulation of a GATA factor. *Molecular Cell* 36, 984–995.
- Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* 34, 77–137.
- Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237–1251.
- Lehmann, A.R. (2001). The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev* 15, 15–23.
- Lemarchandel, V., Ghysdael, J., Mignotte, V., Rahuel, C., and Roméo, P.H. (1993). GATA and Ets cis-acting sequences mediate megakaryocyte-specific expression. *Mol Cell Biol* 13, 668–676.
- Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12, 1725–1735.
- Letting, D.L., Chen, Y.-Y., Rakowski, C., Reedy, S., and Blobel, G.A. (2004). Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci USA* 101, 476–481.
- Letting, D.L., Rakowski, C., Weiss, M.J., and Blobel, G.A. (2003). Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Mol Cell Biol* 23, 1334–1340.

- Li, L., Lee, J.Y., Gross, J., Song, S.-H., Dean, A., and Love, P.E. (2010). A requirement for Lim domain binding protein 1 in erythropoiesis. *J Exp Med* 207, 2543–2550.
- Li, Z., Godinho, F.J., Klusmann, J.-H., Garriga-Canut, M., Yu, C., and Orkin, S.H. (2005). Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet* 37, 613–619.
- Liew, C.K., Kowalski, K., Fox, A.H., Newton, A., Sharpe, B.K., Crossley, M., and Mackay, J.P. (2000). Solution structures of two CCHC zinc fingers from the FOG family protein U-shaped that mediate protein-protein interactions. *Structure* 8, 1157–1166.
- Liew, C.K., Simpson, R.J.Y., Kwan, A.H.Y., Crofts, L.A., Loughlin, F.E., Matthews, J.M., Crossley, M., and Mackay, J.P. (2005). Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. *Proc Natl Acad Sci USA* 102, 583–588.
- Lomvardas, S., and Thanos, D. (2001). Nucleosome sliding via TBP DNA binding in vivo. *Cell* 106, 685–696.
- Lourenço, D., Brauner, R., Rybczynska, M., Nihoul-Fékété, C., McElreavey, K., and Bashamboo, A. (2011). Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proc Natl Acad Sci USA* 108, 1597–1602.
- Ludlow, L.B., Schick, B.P., Budarf, M.L., Driscoll, D.A., Zachai, E.H., Cohen, A., and Konkle, B.A. (1996). Identification of a mutation in a GATA binding site of the platelet glycoprotein Ibbeta promoter resulting in the Bernard-Soulier syndrome. *J Biol Chem* 271, 22076–22080.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Malinge, S., Izraeli, S., and Crispino, J.D. (2009). Insights into the manifestations, outcomes, and mechanisms of leukemogenesis in Down syndrome. *Blood* 113, 2619–2628.
- Mar, B.G., Bullinger, L., Basu, E., Schlis, K., Silverman, L.B., Döhner, K., and Armstrong, S.A. (2012). Sequencing histone-modifying enzymes identifies UTX mutations in acute lymphoblastic leukemia. *Leukemia* 26, 1881–1883.
- Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* 271, 27176–27183.
- Martin, D.I., and Orkin, S.H. (1990). Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev* 4, 1886–1898.
- Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu Rev Genom Human Genet* 7, 29–59.
- Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem* 255, 11992–11996.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195.
- McKercher, S.R., Torbett, B.E., Anderson, K.L., Henkel, G.W., Vestal, D.J., Baribault, H., Klemsz,

- M., Feeney, A.J., Wu, G.E., Paige, C.J., et al. (1996). Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *Embo J* 15, 5647–5658.
- Mehaffey, M.G., Newton, A.L., Gandhi, M.J., Crossley, M., and Drachman, J.G. (2001). X-linked thrombocytopenia caused by a novel mutation of GATA-1. *Blood* 98, 2681–2688.
- Meier, N., Krpic, S., Rodriguez, P., Strouboulis, J., Monti, M., Krijgsveld, J., Gering, M., Patient, R., Hostert, A., and Grosveld, F. (2006). Novel binding partners of Ldb1 are required for haematopoietic development. *Development* 133, 4913–4923.
- Merika, M., and Orkin, S.H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLF. *Mol Cell Biol* 15, 2437–2447.
- Miccio, A., Wang, Y., Hong, W., Gregory, G.D., Wang, H., Yu, X., Choi, J.K., Shelat, S., Tong, W., Poncz, M., et al. (2010). NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. *Embo J* 29, 442–456.
- Michaud, J., Wu, F., Osato, M., Cottles, G.M., Yanagida, M., Asou, N., Shigesada, K., Ito, Y., Benson, K.F., Raskind, W.H., et al. (2002). In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis. *Blood* 99, 1364–1372.
- Mikkola, H.K.A., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., and Orkin, S.H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* 421, 547–551.
- Miller, I.J., and Bieker, J.J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol Cell Biol* 13, 2776–2786.
- Mills, J.A., Wang, K., Paluru, P., Ying, L., Lu, L., Galvão, A.M., Xu, D., Yao, Y., Sullivan, S.K., Sullivan, L.M., et al. (2013). Clonal genetic and hematopoietic heterogeneity among human-induced pluripotent stem cell lines. *Blood* 122, 2047–2051.
- Montes, M., Becerra, S., Sánchez-Álvarez, M., and Suñé, C. (2012). Functional coupling of transcription and splicing. *Gene* 501, 104–117.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34, 267–273.
- Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther* 7, 1063–1066.
- Muntean, A.G., and Crispino, J.D. (2005). Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood* 106, 1223–1231.
- Murawska, M., and Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle. *Transcription* 2, 244–253.

- Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nat Genet* 39, 1507–1511.
- Nerlov, C., Querfurth, E., Kulesa, H., and Graf, T. (2000). GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood* 95, 2543–2551.
- Nesbit, M.A., Bowl, M.R., Harding, B., Ali, A., Ayala, A., Crowe, C., Dobbie, A., Hampson, G., Holdaway, I., Levine, M.A., et al. (2004). Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. *J Biol Chem* 279, 22624–22634.
- Newton, A., Mackay, J., and Crossley, M. (2001). The N-terminal zinc finger of the erythroid transcription factor GATA-1 binds GATC motifs in DNA. *J Biol Chem* 276, 35794–35801.
- Nichols, K.E., Crispino, J.D., Poncz, M., White, J.G., Orkin, S.H., Maris, J.M., and Weiss, M.J. (2000). Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* 24, 266–270.
- Nicodeme, E., Jeffrey, K.L., Schaefer, U., Beinke, S., Dewell, S., Chung, C.-W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H., et al. (2010). Suppression of inflammation by a synthetic histone mimic. *Nature* 468, 1119–1123.
- Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309.
- Nucifora, G., and Rowley, J.D. (1995). AML1 and the 8;21 and 3;21 translocations in acute and chronic myeloid leukemia. *Blood* 86, 1–14.
- Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* 375, 316–318.
- Nutt, S.L., Metcalf, D., D'Amico, A., Polli, M., and Wu, L. (2005). Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. *J Exp Med* 201, 221–231.
- Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science* 183, 330–332.
- Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J., and Gronenborn, A.M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* 261, 438–446.
- Ono, Y., Fukuhara, N., and Yoshie, O. (1998). TAL1 and LIM-only proteins synergistically induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia by acting as cofactors for GATA3. *Mol Cell Biol* 18, 6939–6950.
- Orkin, S.H. (1992). GATA-binding transcription factors in hematopoietic cells. *Blood* 80, 575–581.
- Orkin, S.H., Harosi, F.I., and Leder, P. (1975). Differentiation in erythroleukemic cells and their somatic hybrids. *Proc Natl Acad Sci USA* 72, 98–102.
- Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132, 631–644.

- Orozco, I.J., Kim, S.J., and Martinson, H.G. (2002). The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. *J Biol Chem* 277, 42899–42911.
- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev* 10, 2657–2683.
- Orphanides, G., Leroy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* 92, 105–116.
- Osada, H., Grutz, G., Axelson, H., Forster, A., and Rabbitts, T.H. (1995). Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci USA* 92, 9585–9589.
- Osborne, A., Tschickardt, M., and Blanck, G. (1997). Retinoblastoma protein expression facilitates chromatin remodeling at the HLA-DRA promoter. *Nucleic Acids Res* 25, 5095–5102.
- Ostergaard, P., Simpson, M.A., Connell, F.C., Steward, C.G., Brice, G., Woollard, W.J., Dafou, D., Kilo, T., Smithson, S., Lunt, P., et al. (2011). Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* 43, 929–931.
- Overington, J.P., Al-Lazikani, B., and Hopkins, A.L. (2006). How many drug targets are there? *Nat Rev Drug Discov* 5, 993–996.
- Pal, S., Cantor, A.B., Johnson, K.D., Moran, T.B., Boyer, M.E., Orkin, S.H., and Bresnick, E.H. (2004). Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci USA* 101, 980–985.
- Pang, L., MJ, W., and Poncz, M. (2005). Megakaryocyte biology and related disorders. *J Clin Invest* 115, 3332–3338.
- Pasquet, M., Bellanné-Chantelot, C., Tavitian, S., Prade, N., Beaupain, B., Larochelle, O., Petit, A., Rohrlach, P., Ferrand, C., Van Den Neste, E., et al. (2013). High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* 121, 822–829.
- Perkins, A.C., Sharpe, A.H., and Orkin, S.H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* 375, 318–322.
- Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H., and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. *Development* 121, 163–172.
- Phillips, J.D., Steensma, D.P., Pulsipher, M.A., Spangrude, G.J., and Kushner, J.P. (2007). Congenital erythropoietic porphyria due to a mutation in GATA1: the first trans-acting mutation causative for a human porphyria. *Blood* 109, 2618–2621.
- Pilon, A.M., Ajay, S.S., Kumar, S.A., Steiner, L.A., Cherukuri, P.F., Wincovitch, S., Anderson, S.M., NISC Comparative Sequencing Center, Mullikin, J.C., Gallagher, P.G., et al. (2011). Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood* 118, e139–e148.

- Pilon, A.M., Arcasoy, M.O., Dressman, H.K., Vayda, S.E., Maksimova, Y.D., Sangerman, J.I., Gallagher, P.G., and Bodine, D.M. (2008). Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol Cell Biol* 28, 7394–7401.
- Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* 86, 47–57.
- Rahman, S., Sowa, M.E., Ottinger, M., Smith, J.A., Shi, Y., Harper, J.W., and Howley, P.M. (2011). The Brd4 Extraterminal Domain confers transcription activation independent of pTEFb by recruiting multiple proteins including NSD3. *Mol Cell Biol* 31, 2641–2652.
- Rando, O.J., and Chang, H.Y. (2009). Genome-wide views of chromatin structure. *Annu Rev Biochem* 78, 245–271.
- Raskind, W.H., Niakan, K.K., Wolff, J., Matsushita, M., Vaughan, T., Stamatoyannopoulos, G., Watanabe, C., Rios, J., and Ochs, H.D. (2000). Mapping of a syndrome of X-linked thrombocytopenia with thalassemia to band Xp11-12: further evidence of genetic heterogeneity of X-linked thrombocytopenia. *Blood* 95, 2262–2268.
- Reamon-Buettner, S.M., and Borlak, J. (2005). GATA4 zinc finger mutations as a molecular rationale for septation defects of the human heart. *J Med Genet* 42, e32.
- Reines, D., Chamberlin, M.J., and Kane, C.M. (1989). Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. *J Biol Chem* 264, 10799–10809.
- Rekhtman, N., Radparvar, F., Evans, T., and Skoultschi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev* 13, 1398–1411.
- Rekhtman, N., Choe, K.S., Matushansky, I., Murray, S., Stopka, T., and Skoultschi, A.I. (2003). PU.1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation. *Mol Cell Biol* 23, 7460–7474.
- Rieger, M.A., and Schroeder, T. (2012). Hematopoiesis. *Cold Spring Harbor Perspectives in Biology* 4, a008250–a008250.
- Risheg, H., Graham, J.M., Clark, R.D., Rogers, R.C., Opitz, J.M., Moeschler, J.B., Peiffer, A.P., May, M., Joseph, S.M., Jones, J.R., et al. (2007). A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat Genet* 39, 451–453.
- Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J.R., Vyas, P., De Boer, E., Grosveld, F., and Strouboulis, J. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo J* 24, 2354–2366.
- Ryan, D.P., Sunde, M., Kwan, A.H.Y., Marianayagam, N.J., Nancarrow, A.L., Vanden Hoven, R.N., Thompson, L.S., Baca, M., Mackay, J.P., Visvader, J.E., et al. (2006). Identification of the key LMO2-binding determinants on Ldb1. *J Mol Biol* 359, 66–75.
- Rylski, M., Welch, J.J., Chen, Y.-Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A., and Weiss, M.J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. *Mol Cell Biol*

Biol 23, 5031–5042.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* 7, 437–447.

Saleque, S., Cameron, S., and Orkin, S.H. (2002). The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev* 16, 301–306.

Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.-A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., et al. (2012). Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest* 122, 2439–2443.

Schott, J.J., Benson, D.W., Basson, C.T., Pease, W., Silberbach, G.M., Moak, J.P., Maron, B.J., Seidman, C.E., and Seidman, J.G. (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 281, 108–111.

Schuh, A.H., Tipping, A.J., Clark, A.J., Hamlett, I., Guyot, B., Iborra, F.J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P., et al. (2005). ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. *Mol Cell Biol* 25, 10235–10250.

Schwartz, C.E., Tarpey, P.S., Lubs, H.A., Verloes, A., May, M.M., Risheg, H., Friez, M.J., Futreal, P.A., Edkins, S., Teague, J., et al. (2007). The original Lujan syndrome family has a novel missense mutation (p.N1007S) in the MED12 gene. *J Med Genet* 44, 472–477.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 265, 1573–1577.

Shandilya, J., and Roberts, S.G.E. (2012). The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *BBA - Gene Regulatory Mechanisms* 1819, 391–400.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442, 96–99.

Shilatifard, A., Lane, W.S., Jackson, K.W., Conaway, R.C., and Conaway, J.W. (1996). An RNA polymerase II elongation factor encoded by the human ELL gene. *Science* 271, 1873–1876.

Shimizu, R., Takahashi, S., Ohneda, K., Engel, J.D., and Yamamoto, M. (2001). In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *Embo J* 20, 5250–5260.

Shimizu, R., Kobayashi, E., Engel, J.D., and Yamamoto, M. (2009). Induction of hyperproliferative fetal megakaryopoiesis by an N-terminally truncated GATA1 mutant. *Genes Cells* 14, 1119–1131.

Shirihai, O.S., Gregory, T., Yu, C., Orkin, S.H., and Weiss, M.J. (2000). ABC-me: a novel mitochondrial transporter induced by GATA-1 during erythroid differentiation. *Embo J* 19, 2492–2502.

Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo J* 16, 3965–3973.

- Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844–847.
- Siatecka, M., and Bieker, J.J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood* **118**, 2044–2054.
- Singleton, B.K., Roxby, D.J., Stirling, J.W., Spring, F.A., Wilson, C., Poole, J., and Anstee, D.J. (2013). A novel GATA1 mutation (Stop414Arg) in a family with the rare X-linked blood group Lu(a-b-) phenotype and mild macrothrombocytic thrombocytopenia. *Br J Haematol* **161**, 139–142.
- Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146**, 318–331.
- Soler, E., Andrieu-Soler, C., De Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.-J., Stevens, M., Kockx, C., Van Ijcken, W., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24**, 277–289.
- Song, S.-H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Molecular Cell* **28**, 810–822.
- Song, W.J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* **23**, 166–175.
- Southwood, C.M., Downs, K.M., and Bieker, J.J. (1996). Erythroid Krüppel-like factor exhibits an early and sequentially localized pattern of expression during mammalian erythroid ontogeny. *Dev Dyn* **206**, 248–259.
- Stachura, D.L., Chou, S.T., and Weiss, M.J. (2006). Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. *Blood* **107**, 87–97.
- Stopka, T., Amanatullah, D.F., Papetti, M., and Skoultschi, A.I. (2005). PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *Embo J* **24**, 3712–3723.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550.
- Sun, X.J., Wei, J., Wu, X.-Y., Hu, M., Wang, L., Wang, H.-H., Zhang, Q.-H., S.-J, C., Huang, Q.-H., and Z, C. (2005). Identification and Characterization of a Novel Human Histone H3 Lysine 36-specific Methyltransferase. *J Biol Chem* **280**, 35261–35271.
- Tallack, M.R., Whittington, T., Yuen, W.S., Wainwright, E.N., Keys, J.R., Gardiner, B.B., Nourbakhsh, E., Cloonan, N., Grimmond, S.M., Bailey, T.L., et al. (2010). A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* **20**, 1052–1063.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation

as a new type of histone modification. *Cell* **146**, 1016–1028.

Tijssen, M.R., Cvejic, A., Joshi, A., Hannah, R.L., Ferreira, R., Forrai, A., Bellissimo, D.C., Oram, S.H., Smethurst, P.A., Wilson, N.K., et al. (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Developmental Cell* **20**, 597–609.

Tjian, R. (1995). Molecular machines that control genes. *Sci Am* **272**, 54–61.

Trainor, C.D., Omichinski, J.G., Vandergon, T.L., Gronenborn, A.M., Clore, G.M., and Felsenfeld, G. (1996). A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol Cell Biol* **16**, 2238–2247.

Tripic, T., Deng, W., Cheng, Y., Zhang, Y., Vakoc, C.R., Gregory, G.D., Hardison, R.C., and Blobel, G.A. (2009). SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* **113**, 2191–2201.

Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G., and Orkin, S.H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339**, 446–451.

Tsang, A.P., Fujiwara, Y., Hom, D.B., and Orkin, S.H. (1998). Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* **12**, 1176–1188.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M., and Orkin, S.H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**, 109–119.

Tubman, V.N., Levine, J.E., Campagna, D.R., Monahan-Earley, R., Dvorak, A.M., Neufeld, E.J., and Fleming, M.D. (2007). X-linked gray platelet syndrome due to a GATA1 Arg216Gln mutation. *Blood* **109**, 3297–3299.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular Cell* **17**, 453–462.

Varley, J.M. (2003). Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat* **21**, 313–320.

Villeval, J.L., Cohen-Solal, K., Tulliez, M., Giraudier, S., Guichard, J., Burstein, S.A., Cramer, E.M., Vainchenker, W., and Wendling, F. (1997). High thrombopoietin production by hematopoietic cells induces a fatal myeloproliferative syndrome in mice. *Blood* **90**, 4369–4383.

Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H., and Shivdasani, R.A. (1999). Consequences of GATA-1 deficiency in megakaryocytes and platelets. *Blood* **93**, 2867–2875.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**, 343–356.

- Wadman, I.A., Osada, H., Grütz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo J* 16, 3145–3157.
- Wall, L., deBoer, E., and Grosveld, F. (1988). The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev* 2, 1089–1100.
- Wang, X., Crispino, J.D., Letting, D.L., Nakazawa, M., Poncz, M., and Blobel, G.A. (2002). Control of megakaryocyte-specific gene expression by GATA-1 and FOG-1: role of Ets transcription factors. *Embo J* 21, 5225–5234.
- Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M., and Crispino, J.D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet* 32, 148–152.
- Wei, D., Bao, H., Liu, X.-Y., Zhou, N., Wang, Q., Li, R.-G., Xu, Y.-J., and Yang, Y.-Q. (2013). GATA5 loss-of-function mutations underlie tetralogy of fallot. *Int J Med Sci* 10, 34–42.
- Weiss, M.J., and Orkin, S.H. (1995). Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci USA* 92, 9623–9627.
- Weiss, M.J., Keller, G., and Orkin, S.H. (1994). Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* 8, 1184–1197.
- Weiss, M.J., Yu, C., and Orkin, S.H. (1997). Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol* 17, 1642–1651.
- Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A., and Weiss, M.J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* 104, 3136–3147.
- Wilber, A., Tschulena, U., Hargrove, P.W., Kim, Y.-S., Persons, D.A., Barbas, C.F., and Nienhuis, A.W. (2010). A zinc-finger transcriptional activator designed to interact with the gamma-globin gene promoters enhances fetal hemoglobin production in primary human adult erythroblasts. *Blood* 115, 3033–3041.
- Wilkinson-White, L., Gamsjaeger, R., Dastmalchi, S., Wienert, B., Stokes, P.H., Crossley, M., Mackay, J.P., and Matthews, J.M. (2011). Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. *Proc Natl Acad Sci USA* 108, 14443–14448.
- Wilson, B.G., and Roberts, C.W.M. (2011). SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* 11, 481–492.
- Woodcock, C.L., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Exp Cell Res* 97, 101–110.
- Wu, S.-Y., and Chiang, C.-M. (2007). The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J Biol Chem* 282, 13141–13145.

- Xu, Z., Meng, X., Cai, Y., Liang, H., Nagarajan, L., and Brandt, S.J. (2007). Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins. *Genes Dev* 21, 942–955.
- Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular Cell* 21, 227–237.
- Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc Natl Acad Sci USA* 95, 3890–3895.
- Yamagata, K., Furuta, H., Oda, N., Kaisaki, P.J., Menzel, S., Cox, N.J., Fajans, S.S., Signorini, S., Stoffel, M., and Bell, G.I. (1996). Mutations in the hepatocyte nuclear factor-4 α gene in maturity-onset diabetes of the young (MODY1). *Nature* 384, 458–460.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 97, 41–51.
- Yang, Z., He, N., and Zhou, Q. (2008). Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. *Mol Cell Biol* 28, 967–976.
- Yang, Z., Yik, J.H.N., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell* 19, 535–545.
- Yu, C., Niakan, K.K., Matsushita, M., Stamatoyannopoulos, G., Orkin, S.H., and Raskind, W.H. (2002). X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood* 100, 2040–2045.
- Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Molecular Cell* 36, 682–695.
- Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. *Cell Res* 21, 564–578.
- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39, 1512–1516.
- Zeng, P.-Y., Vakoc, C.R., Chen, Z.-C., Blobel, G.A., and Berger, S.L. (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *BioTechniques* 41, 694–696–698.
- Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H.S., Auron, P.E., Tenen, D.G., and Sun, Z. (1999). Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci USA* 96, 8705–8710.
- Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood* 96, 2641–2648.

Zhang, S.-Y., Boisson-Dupuis, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., Picard, C., Abel, L., Jouanguy, E., and Casanova, J.-L. (2008). Inborn errors of interferon (IFN)-mediated immunity in humans: insights into the respective roles of IFN-alpha/beta, IFN-gamma, and IFN-lambda in host defense. *Immunol Rev* 226, 29–40.

Zhao, R., Nakamura, T., Fu, Y., Lazar, Z., and Spector, D.L. (2011). Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nat Cell Biol* 13, 1295-1304.

Zhou, D., Liu, K., Sun, C.-W., Pawlik, K.M., and Townes, T.M. (2010). KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. *Nat Genet* 42, 742–744.

Zon, L.I., Tsai, S.F., Burgess, S., Matsudaira, P., Bruns, G.A., and Orkin, S.H. (1990). The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. *Proc Natl Acad Sci USA* 87, 668–672.