Transcription Factor Networks in Erythroid Cell and Megakaryocyte Development

Louis C. Doré and John D. Crispino

Division of Hematology/Oncology, Northwestern University, Chicago, IL 60611

1Corresponding author:
John Crispino, PhD
Northwestern University
303 East Superior Street
Lurie 5-113
Chicago, IL 60611
312-503-1504 (phone)
312-503-0189 (fax)
j-crispino@northwestern.edu

Review

Running title: Transcription factor complexes in hematopoiesis

Keywords: Hematopoiesis, GATA, ChIP-Seq, KLF1, SCL
Erythroid cells and megakaryocytes are derived from a common precursor, the megakaryocyte-erythroid progenitor (MEP). Although these two closely related hematopoietic cell types share many transcription factors, there are several key differences in their regulatory networks that lead to differential gene expression downstream of the MEP. With the advent of next generation sequencing and our ability to precisely define transcription factor chromatin occupancy in vivo on a global scale, we are much closer to understanding how these two lineages are specified and in general how transcription factor complexes govern hematopoiesis.
Due to the relative accessibility of mature and progenitor cell types, their genetic tractability, and the availability of various functional assays, the hematopoietic system has long been used as a paradigm to study stem cells, lineage decisions, and gene regulation mediated by tissue-specific transcription factors. Both the relative and absolute levels of lineage transcription factors control fate decision and gene expression in specialized cell types. As hematopoietic cells differentiate from stem cells to the mature lineages, they gradually become more committed to their ultimate lineage, therein losing differentiation potential and gaining more specialized functionalities and differentiation focus. In the megakaryocyte and erythroid lineages, the final stage of non-committed progenitor cell type is thought to be a common megakaryocyte-erythroid progenitor (MEP). Progenitor cells such as the MEP are exceedingly rare and difficult to isolate due to their transient nature, although several lines of evidence support the existence of such a cell type in vivo. First, Vainchenker and colleagues identified a human CD34+CD38low cell population that was capable of giving rise to colonies that contain both erythroid and megakaryocytic cells.1 Second, in mice recovering from phenylhydrazine-induced anemia, a cell population expressing both erythroid and megakaryocytic markers can be isolated from spleens and they form colonies containing both cell types.2 Third, Weissman and colleagues successfully isolated a CD34 FcRlo fraction from mouse bone marrow that was capable of generating cells of either megakaryocytic or erythroid phenotype in single cell differentiation experiments.3 Finally, from Gata1 null murine fetal livers, Stachura et al isolated developmentally arrested cells which, when provided GATA-1, were capable of differentiating to both erythroid and megakaryocytic cells.4 To completely understand the maturation program of megakaryocytes and erythrocytes (and the diseases that arise when differentiation goes awry), it is imperative that we define and characterize the transcription factors that
specify and control the differences between the mature lineages as well as those that regulate their similarities and their common progenitor, the MEP.

In the cells that comprise the myeloid compartment of mammalian blood systems, transcriptional regulatory complexes are nucleated by factors that bind GATA or ETS motifs and these factors work together or antagonistically to drive differentiation. Indeed, the specification and differentiation of erythroid cells and megakaryocytes from the MEP is governed by coordinated regulation of a precise balance of members of several classes of factors, including GATA-binding transcription factors (GATA-1 and GATA-2), ETS factors (FLI1 and GABPα), Krüppel-containing factors (KLF1 and LRF), basic helix-loop-helix factors (SCL), multiple adaptors (FOG-1 and LDB1) and even microRNAs (miR-150 and miR-451) (Figure 1). A large number of recent studies have leveraged chromatin immunoprecipitation and next generation sequencing technologies to define the full repertoire of genomic loci occupied by hematopoietic transcription factors. In this review, we will discuss the core factors relevant to erythroid and megakaryocytic differentiation, highlight recent discoveries related to transcription factor occupancy in hematopoietic cells, and provide insights into the future of the field and emerging technologies for the study of transcriptional regulation.

GATA-2

Mice lacking the transcription factor GATA-2 die around embryonic day 10 due to severe anemia. Mouse chimera studies and gain- and loss-of-function experiments utilizing in vitro differentiated mouse ES cells demonstrate a clear requirement for GATA-2 in the proliferation, survival, and/or maintenance of early hematopoietic progenitor cells. Moreover, enforced expression of GATA-2 in hematopoietic bone marrow progenitors severely reduces their capacity to form colonies or contribute to reconstitution of a
transplanted mouse. Interestingly, haploinsufficiency for GATA-2 also diminishes colony formation and in vivo proliferation of hematopoietic stem and progenitor cells in both the bone marrow and the AGM. Together, these data show that a precise dosage of GATA-2 is critical for early hematopoiesis and GATA-2 downregulation is necessary for the initiation of differentiation.

GATA-2 Chromatin Occupancy in Hematopoietic Stem & Progenitor Cells

Two recent studies have explored the genome-wide occupancy pattern of GATA-2 in hematopoietic stem or progenitor cells. The first was a comprehensive genome-wide survey of 10 hematopoietic transcription factors in a murine hematopoietic progenitor cell line, called HPC-7. Here, Göttgens and colleagues show that GATA-2 binding significantly overlaps with that of SCL, LYL1, LMO2, and RUNX1. Additionally, these five factors are commonly found at sites that are also occupied by ERG and FLI1. Collectively, these seven factors comprise the “HSPC heptad” and the targets of this heptad are likely to be critical effectors of hematopoietic development and function.

The second study to examine GATA-2 occupancy in blood progenitor cells used primary lineage-negative bone marrow cells from mice. This study aimed to define the transcriptional program of the hematopoietic adaptor protein LDB1; in doing so, the investigators interrogated LDB1 occupancy as well as the occupancy of the complex components GATA-2 and SCL. LDB1 occupies many critical hematopoietic genes and nearly all of those genes are also occupied by GATA-2 and/or SCL, suggesting that LDB1 is an important component of the transcriptional regulatory network in hematopoietic stem and progenitor cells.

GATA-2 In Maturing Erythrocytes and Megakaryocytes

Doré and Crispino
In addition to its roles in hematopoietic stem cell maintenance, GATA-2 is also expressed in erythroblasts and megakaryocytes. In a multipotent human leukemia cell line (K562), GATA-2 overexpression drives megakaryocytic differentiation at the expense of erythroid differentiation. In maturing erythroid cells, GATA-2 expression is rapidly downregulated upon the activation of GATA-1 – which is itself a GATA-2 target gene. Enforced expression of GATA-2 during erythroid development impairs differentiation, presumably by blocking GATA-1-mediated downregulation of the proliferation and maturation programs. In committed, undifferentiated red blood cells GATA-2 binds to the β-globin locus and activates low-level expression of the appropriate globin gene. As GATA-1 protein levels increase, GATA-1 replaces GATA-2 at the β-globin locus and induces massive upregulation of hemoglobin gene expression. This pattern of GATA occupancy switching during erythroid development is also seen at other GATA target genes, including Gata2, Kit, Lyl1, Sfpi1 (which encodes PU.1), and the miR144/451 microRNA cluster. In the G1E erythroid cell line, GATA-2 and SCL co-occupy the GATA switch sites in the Gata2 and Kit loci while the genes are active. When GATA-1 activity is restored in these cells, GATA-1 replaces GATA-2 at the switch site and SCL occupancy is lost, suggesting that SCL is cooperating with GATA-2 to drive transcription of these key players in early erythropoiesis. Interestingly, GATA-2 also activates mast cell genes in G1E cells, where GATA-2 levels are elevated. Following the GATA switch, GATA-1 downregulates GATA-2, replaces GATA-2 at these mast cell gene promoters, and represses mast cell gene expression, thereby reinforcing the erythroid lineage decision.

Using ChIP-PCR at dozens of GATA consensus sites in G1E cells, Wozniak et al showed that GATA-2 binds to only a small proportion of DNA elements containing
WGATAR motifs in complex with E-box motifs (SCL recognition sites). Moreover, the presence of an E-box near a WGATAR does not significantly enhance the probably that GATA-2 will occupy a site. However, a high percentage of sites which were bound by GATA-2 were marked by a specific epigenetic signature – specifically, these sites were also occupied by SCL, associated with acetylated histones H3 & H4 that were dimethylated at H3K4 and H3K36, and devoid of H3K9 trimethylation. Together, these findings begin to establish a model through which GATA-2 chromatin site selection may be predicted.

Interaction with FOG-1

GATA-2 also associates with the hematopoietic co-factor Friend of GATA-1 (FOG-1), which was initially identified by two-hybrid screening as a binding partner of GATA-1. In early erythroid cells, FOG-1 association with GATA-2 precedes the GATA switch and may facilitate GATA-1 occupancy. Thus, when FOG-1 is absent or when the GATA-1:FOG-1 interaction is perturbed, GATA-2 is not effectively silenced. In the megakaryocytic progenitor line, G1ME, GATA-2 and FOG-1 occupy the Sfpi1 gene promoter and repress its expression. Upon differentiation by restoration of GATA-1, GATA-1 replaces GATA-2 at the promoter and further represses Sfpi1 expression. This active, graded repression of PU.1 by GATA factors is necessary for maintaining the megakaryocytic identity of these progenitor cells, as loss of GATA-2 leads to an upregulation of PU.1 levels and a trans-differentiation of G1ME cells to functional macrophage-like cells. Accordingly, ectopic expression of GATA-2 in a macrophage-directed in vitro ES cell differentiation system blocked macrophage differentiation and redirected output to other hematopoietic lineages.

Other GATA-2 Target Genes in Erythromegakaryocytic Development
Using loss-of-function and chromatin immunoprecipitation experiments in G1ME cells, other GATA-2 target genes have been identified. In addition to Sfpi1, GATA-2 direct targets include other myeloid lineage transcription factors, such as Mpo, Cebpa, and Hhex, and cell cycle regulators such as E2f2, Skp2, Cdkn1a (p21), and Cdkn1b (p27).\textsuperscript{28} ChIP-Seq experiments in K562 cells identified thousands of GATA-2 binding sites, including several binding sites upstream of the CBFA2T3 promoter.\textsuperscript{30} CBFA2T3 encodes the ETO2 transcription factor, a co-repressor known to be in a complex with SCL.\textsuperscript{31} In K562 and G1E cells, GATA-2 binds and activates the CBFA2T3 gene. As differentiation proceeds, ETO2 binds to and represses its own promoter. The derepression of ETO2 target genes, such as Hbb, Eraf, and Slc4a1 encourages erythroid maturation and allows for the switch from a GATA-2-regulated transcriptional program to one driven by GATA-1.\textsuperscript{30} The full repertoire of GATA-2 target genes throughout differentiation remains unknown, but it will be instructive to determine how these targets differ in hematopoietic stem cells, erythroblasts, and megakaryocytes.

**GATA-1**

GATA-1 is perhaps the best-studied hematopoietic transcription factor. GATA-1 deficient mice die around embryonic day 10 due to severe anemia, despite ~50-fold increased expression of GATA-2 in GATA-1 null proerythroblasts.\textsuperscript{32-34} GATA-1 is expressed primarily in mature cell types and is required for terminal maturation of several blood lineages, including erythrocytes, megakaryocytes, eosinophils, and mast cells.\textsuperscript{4,34-38}

**GATA-1 and FOG-1 Chromatin Occupancy**

In red blood cells, GATA-1 represses nearly as many genes as it activates. One of its primary target genes is also its major binding partner, FOG-1, which functions with GATA-1 in a feed-forward regulatory circuit to activate β-globin expression during
erythropoiesis. FOG-1 interacts with the N-terminal zinc finger of GATA-1 and functions as a coactivator in reporter assays. Despite the fact that FOG-1 does not bind chromatin directly, the absence of FOG-1 can prevent GATA-1 from binding normally to its target sites. At the Gata2 locus, GATA-1 is capable of binding even when its interaction with FOG-1 is disrupted by a point mutation in GATA-1. However, because FOG-1 cannot be properly recruited, GATA-1 fails to recruit enzymes that deacetylate histones H3 and H4 and repress Gata2 gene expression. Thus, in some contexts, FOG-1 serves as a transcriptional co-repressor (see The GATA-1:FOG-1:NuRD Regulatory Axis).

FOG-1 is responsible for more than just facilitating GATA-1 chromatin occupancy. Indeed, following GATA-1 binding at the β-globin locus, GATA-1 and FOG-1 cooperate to induce the formation of a chromatin loop between a distal enhancer element, called the locus control region, and the Hbb gene promoter that is critical for proper globin expression in maturing erythrocytes. At the Kit locus, GATA-1 and FOG-1 are also responsible for the rearrangement of a chromatin loop structure coincident with down-regulation of c-Kit expression in a model of maturing red blood cells. Here, GATA-2 and FOG-1 nucleate a chromatin loop between a distal enhancer region (-114kb site) and the Kit promoter in undifferentiated G1E cells, which are developmentally arrested, committed erythroid progenitors that proliferate. After restoration of GATA-1, the loop structure is rearranged such that GATA-1 and FOG-1 are integral players in a chromatin loop between the promoter and the +58kb region in the transcribed region of the Kit locus. In the absence of the GATA-1:FOG-1 interaction, this repressive loop does not form and Kit expression is not downregulated. Moreover, experiments utilizing mice harboring a BRG1 hypomorphic mutant suggest that the chromatin remodeler BRG1 is
critical for the formation of GATA-1-mediated chromatin loops, at least at the \( Hbb \) locus.\(^{43} \)

The GATA-1:FOG-1:NuRD Regulatory Axis

To determine the mechanism by which FOG-1 represses GATA-1 target genes, such as \( Gata2 \), \( Myc \), and \( Kit \),\(^{44-46} \) FOG-1 was used to immunoprecipitate partner proteins from murine erythroleukemia (MEL) cell extracts. The N-terminus of FOG-1 interacts with a complex of transcriptional repressors known as the nucleosome remodeling and histone deacetylase (NuRD) complex.\(^{47,48} \) The interaction between FOG-1 and NuRD components is critical for GATA-1 repression of \( Gata2 \) and \( Kit \), and a variety of mast cell genes.\(^{47} \) NuRD also associates with FOG-1 at sites where GATA-1 activates transcription.\(^{49} \) Given the deacetylase activity of the NuRD complex, this was unexpected. The precise mechanism through which NuRD activates GATA-1/FOG-1 targets remains unclear. It is clear, however, that the FOG-1:NuRD interaction is critical for the maintenance of lineage fidelity within the erythroid and megakaryocytic compartments.\(^{24,50} \) Specifically, a number of mast cell genes, including carboxypeptidases (\( Cpa3, Cpd \)), mast cell proteases (\( Mcpt2, Mcpt6 \)), and FC receptors (\( Fcer1a, Fcer1b \)) are direct GATA target genes. In G1E cells, in the absence of GATA-1, artificially high levels of GATA-2 activate these transcripts; FOG-1 and NuRD also occupy the active genes. Upon induction of differentiation, GATA-1 replaces GATA-2 while FOG-1 and NuRD remain. GATA-1 binding leads to a FOG-1:NuRD-dependent reduction in histone acetylation and gene expression.\(^{24} \) Taken together, these papers show that a disruption in the GATA-1:FOG-1:NuRD axis of regulation leads to inappropriate lineage gene expression and loss of erythromegakaryocytic identity in a manner that resembles GATA-2 loss in committed GATA-1 null megakaryocyte-erythroid progenitors.\(^{28} \)
LRF Association with GATA-1

In MEL cells, GATA-1 ChIP-Seq experiments identified multiple coinciding transcription factor binding motifs in GATA-1-occupied regions. One of these motifs was a match for the Poz and Krüppel related factor, LRF (for Leukemia/lymphoma Related Factor; formerly known as POKEMON and encoded by the \textit{Zbtb7a} gene).\textsuperscript{51} LRF was previously identified as a GATA-1 target gene whose product is essential for erythroid development in mice.\textsuperscript{52} Intriguingly, using ChIP-PCR experiments, Cantor and colleagues showed that LRF was bound to many GATA-1-bound regions associated with activated gene expression. The finding that LRF binds to GATA-1 target sites suggests the existence of a feed-forward regulatory loop operating through LRF and anchored by GATA-1. A more complete understanding of the molecular function of LRF is likely to manifest within the next few years and LRF ChIP-Seq experiments, which are undoubtedly forthcoming, are sure to be a critical aspect of that understanding.

The GATA-1:SCL Activation Complex

SCL (also known at TAL-1) is a GATA-1 target gene that is required for blood cell development and forms complexes with GATA factors at activating sites.\textsuperscript{21,53,54} SCL is required for the specification of all hematopoietic lineages and SCL-/- mice die around embryonic day 9.5.\textsuperscript{54-57} In contrast to the critical role that SCL has in hematopoietic stem cell generation, it is not required for hematopoietic stem cell survival, multipotency, or long-term repopulating in mice. However, in the absence of SCL, megakaryocytic and erythroid differentiation are severely hampered.\textsuperscript{58} During normal erythroid differentiation, SCL co-occupies active, but not repressive, GATA binding sites along with GATA-1, LMO2, Ldb1, and E2A. At genomic loci where GATA-1 represses transcription, the SCL complex is absent.\textsuperscript{21}
To gain insights into the molecules that contribute to the functions of the various GATA-1:SCL complexes, several groups have begun to identify genome-wide occupancy patterns for individual members of the complexes.

**SCL**

SCL functions as a core member in the GATA-nucleated regulatory complex, which also includes LMO2, LDB1, E2A, and sometimes ETO2. In general, SCL, LMO2, and LDB1 comprise a complex which is primarily activating, while ETO2 serves as a co-repressor.²¹,³⁰,⁵⁹,⁶⁰ Interestingly, mice that harbor a congenital mutation in the basic helix-loop-helix DNA binding domain of SCL (SCLRER) do not all die as early as the complete knockouts, suggesting that SCL DNA binding is not always be direct or that SCL has non-DNA binding functions. Importantly, it appears that DNA binding is dispensable for the HSC specification functions of SCL, but not for SCL functions related to erythroid maturation.⁶¹ Several recent studies have exploited ChIP-chip or ChIP-Seq technology to define occupancy patterns of a number of these components. SCL is coincidentally bound with GATA-1 at a number of GATA-1-activated sites, but not repressive sites, in erythrocytes, megakaryocytes, and mast cells.²¹ Using the SCLRER knockin mouse model, Kassouf et al. recently performed ChIP-Seq to catalog the DNA binding activity of the wild type and mutant SCL.⁶² This approach allowed the authors to discern functional differences between direct and indirect DNA binding by SCL. By integrating their ChIP-Seq data with global gene expression studies, they conclude that direct DNA binding of SCL is necessary for transcriptional activation in red blood cells. Moreover, their evidence in primary Ter119-negative murine fetal liver cells suggests that GATA-1 and SCL function as mutually-stabilizing factors at SCL sites of transcriptional activation.⁶² In HPC-7 cells, SCL regulates a broad network of hematopoietic transcription factors and is
a constituent of the HSPC heptad discussed earlier (see *GATA-2 Chromatin Occupancy in Hematopoietic Stem & Progenitor Cells*).\textsuperscript{14,63}

Other SCL Partner Proteins: LDB1 & LSD1

Another component of the GATA/SCL complex, LDB1, has also been shown to be required for chromatin loop formation at the *Hbb* locus in murine cell lines.\textsuperscript{64} LDB1 null mice do not make red blood cells and die from various morphologic abnormalities and anemia around E9.5.\textsuperscript{65} Recent work has identified a requirement for LDB1 throughout embryonic and adult erythroid and megakaryocytic development. The same study also suggested that LDB1 is a critical downstream effector of the transcriptional activation network of GATA-1.\textsuperscript{66} The LDB1 chromatin occupancy repertoire was also recently interrogated in murine erythroleukemia cells. LDB1 binds to genomic sites primarily as a part of an activating complex and is also believed to have a role in the formation of chromatin loops and long-range interactions between the *Hbb* locus and other LDB1 bound genes on chromosome 7.\textsuperscript{67} However, the purpose of these long-range LDB1 interactions remains unclear.

To identify additional partner proteins of SCL in K562 human erythroleukemia cells, SCL-interacting proteins were isolated by immunoaffinity and identified by mass spectrometry. In addition to the known SCL complex members LDB1 and ETO2, several epigenetic modifier enzymes were discovered. In particular, the histone 3 lysine 4 demethylase LSD1, histone deacetylases HDAC1 & HDAC2, and the corepressor molecule CoREST were prominent hits from this biochemical screen.\textsuperscript{68} Importantly, this complex has been previously shown to be a necessary component of the GFI1-B-mediated transcriptional repression program during hematopoiesis.\textsuperscript{69} Understanding precisely how SCL, LDB1, and other hematopoietic transcription factors identify genomic
sites as targets of activation or repression and subsequently recruit appropriate co-activators or co-repressors, such as LSD1, will be paramount to comprehensively mapping the molecular genetic landscape that drives hematopoietic differentiation.

**KLF1**

The CACCC-binding nuclear factor, KLF1 (also known as EKLF), is the founding member of the mammalian Krüppel-like family of transcription factors. In erythroid cells, *Klf1* is a direct target of the BMP4/Smad pathway and GATA-1. Expression of KLF1 is critical for activation of β-globin transcription and mice lacking KLF1 die in utero. At the β-globin locus, KLF1 recruits a SWI/SNF-related chromatin remodeling complex, E-RC1 (for EKLF coactivator-remodeling complex-1), that is both necessary and sufficient for establishment of DNase hypersensitivity and active transcription of the β-globin gene. KLF1 also interacts with the histone acetyltransferases CBP and p300, which acetylate KLF1 and enhance its transcriptional activity in vitro and in vivo. Through its zinc-finger domain, KLF1 is also capable of interacting with mSin3a and HDAC1, which allow KLF1 to function as a transcriptional repressor in vitro. However, gene expression profiling in fetal liver erythroid progenitors from KLF1−/− mice suggest that KLF1 functions primarily as a transcriptional activator in vivo and has very few, if any, directly repressed erythroid targets. KLF1 is also required for coordination of the three-dimensional chromosome conformation of the β-globin locus and AHSP gene, although the molecular mechanisms by which KLF1 promotes loop formation and chromatin hub assembly remain unknown.

One of the many genes regulated by KLF1 is *Zbtb7a* (which encodes LRF), although the directness of this regulation has yet to be determined. It is interesting to note, however,
that LRF itself occupies the Klf1 promoter along with GATA-1, further complicating the regulatory loop that controls erythroid development.

To gain further insights into the role of KLF1 in erythropoiesis, Perkins and colleagues performed ChIP-Seq for KLF1 in primary murine fetal liver erythroid cells. Using a highly specific peak-calling approach, they identified approximately 1000 genomic binding sites for KLF1. Many of the identified binding sites contain a motif that strongly matches the in vitro predicted consensus binding site (CCNCNCCCN). More than a quarter of the bound sites also contain motifs that match a composite GATA/E-box motif, in agreement with the longstanding notion that KLF1 and GATA1 cooperatively regulate many terminal erythroid targets. Most interestingly, KLF1 directly controls nearly all aspects of the heme synthesis and iron procurement pathway in maturing erythroid cells, establishing a crucial molecular role for KLF1 in the establishment of functional erythrocytes.

KLF1 Activation of BCL11A

The transcription factor BCL11A is a major determinant in fetal-to-adult hemoglobin switching as it directly represses human γ-globin, in concert with SOX6 and GATA-1. Recently, several groups have reported that BCL11A is a direct target of KLF1 in human and mouse erythroid cells. Hereditary persistence of fetal hemoglobin (HPFH) caused by haploinsufficiency for KLF1 in a Maltese family was thus explained by a failure of KLF1 to fully activate BCL11A in adult erythroid cells. Together, these studies suggest that both BCL11A and KLF1 may be attractive targets for therapies to increase fetal hemoglobin expression in patients with β-hemoglobinopathies or thalassemias.

KLF1 in MEPs and Megakaryocytes
In a series of papers around the beginning of 2008, Bieker and colleagues demonstrated a clear role for KLF1 in MEPs and the megakaryocyte-erythroid lineage fate decision. During the mesodermal specification of in vitro differentiated murine embryoid bodies (EB), Klf1 is activated in a GATA-1-independent manner by a GATA-2 and Smad5-nucleated complex. Using a GFP reporter under the control of the Klf1 promoter, they showed that cells which express GFP and not the erythroid-specific marker Ter119 are capable of ultimately differentiating into either megakaryocytic or erythrocytic colonies, suggesting that KLF1 has a role in MEPs. In other work, Frontelo et al. used gain- and loss-of-function studies in EB differentiation systems to demonstrate that enforced KLF1 expression selectively blocks megakaryocyte development (to the benefit of erythroid development) when KLF1 is expressed during a short window prior to megakaryocyte-erythroid lineage choice. Microarray studies coupled with luciferase assays demonstrated that KLF1 represses FLI1 and many other critical megakaryocytic genes. Additionally, Siatecka et al identified a post-translational sumoylation modification which is critical for KLF1 interaction with the NuRD corepressor complex. Mutation of the SUMO conjugation site results in upregulation of KLF1 repression targets, including Fli1, and a loss of KLF1-directed suppression of megakaryopoiesis. Together, these reports demonstrate a role for KLF1 in the MEP and the megakaryocyte-erythroid lineage switch.

Mutual Antagonism between KLF1 & FLI1

Although KLF1 does not directly repress many genes in committed erythroid cells, accumulating evidence that it has a role in repressing alternative lineages during cell fate decision points to an interaction with the megakaryocytic ETS factor FLI-1. In a heterologous cell line, the DNA-binding domain of KLF1 represses FLI-1-mediated transcription of megakaryocytic reporter gene constructs and FLI-1 is capable of...
repressing KLF1 target genes. Based on these findings, Morle and colleagues proposed that, in a bipotential erythromegakaryocytic progenitor cell type, a marginal and probably stochastic increase in KLF1 or FLI1 levels may drive cell fate decision in a manner similar to the balance that exists between PU.1 and GATA-1 in earlier myeloid progenitors.\textsuperscript{94,95} This hypothesis was later supported by gain- and loss-of-function studies using \textit{in vitro} differentiation of murine ES cells,\textsuperscript{92} human CD34+ cells,\textsuperscript{96} and human ES cells.\textsuperscript{97} Additionally, mouse bone marrow reconstituted with KLF1\textsuperscript{-/-} fetal liver cells showed a marked increase in megakaryocytic progenitors and circulating platelets, further suggesting that KLF1 has a role in erythromegakaryocytic cell fate decision \textit{in vivo}.\textsuperscript{98} Moreover, heterozygous Nan mutant mice, which harbor a point mutation in KLF1 that disrupts DNA binding, are severely anemic, but have increased platelet counts.\textsuperscript{99} Conditional deletion of \textit{Fli1} in mouse bone marrow cells leads to increased erythroid output from MEP cells which further supports the existence of a mutually antagonistic relationship between KLF1 and FLI1.\textsuperscript{100}

\textbf{FLI1}

Homozygous loss of functional \textit{Fli1} alleles in mice leads to embryonic lethality due to severe defects in fetal megakaryopoiesis and a high incidence of embryonic hemorrhaging. The latter phenotype is presumably due to coagulation defects secondary to impaired megakaryopoiesis as well as inefficient blood vessel formation, as FLI-1 is also essential for vascular endothelium and hemangioblast specification.\textsuperscript{101-103} Moreover, FLI-1 is a key component, along with GATA-2 and SCL, of a gene regulatory network kernel that controls hematopoietic stem cell specification. In HSCs, these 3 factors comprise a fully connected triad, wherein each factor directly activates the expression of each of the other two factors, leading to a robust and consistently active network module.\textsuperscript{104} In the multipotent HPC-7 line, FLI-1 occupancy closely paralleled genome-
wide occupancy of the closely related ETS factor ERG. Additionally, FLI-1 was identified as a core constituent of a regulatory heptad of transcription factors (ERG, FLI-1, GATA-2, LMO2, LYL1, RUNX1, SCL), including many discussed in this review. Within a population of HPC-7 cells, these factors coincidentally occupy genomic regions associated with hundreds of genes that are enriched for cell death, cell cycle, signaling, and transcriptional control, although there is no evidence that all of these factors are bound to the same span of DNA at the same time within the same cell. Thus, it is possible that (1) different combinations of the seven factors may be binding in different cells within the population, creating the illusion of a large complex of diverse DNA-binding factors that may not exist or (2) different combinations of genes and gene sets are being regulated by these complexes within individual cells throughout the population. As technologies advance, single cell analysis and further biochemical and molecular characterization of the precise components of the regulatory complexes and the genes they control will address these issues.

In megakaryocytic cells, FLI-1 binds to and directly regulates the genes encoding several proteins that are essential for terminal megakaryocytic maturation. Specifically, in the Y10 cell line and primary fetal liver-derived megakaryocytes, GATA-1, FOG-1 and FLI-1 coordinately bind and regulate \( \text{Itga2b} \) (which encodes the CD41 antigen) as well as \( \text{Gp1ba} \) (CD42), \( \text{Gpix} \) (Glycoprotein 9), \( \text{Mpl} \) (Thrombopoietin receptor), and \( \text{Cxcl4} \) (Platelet factor 4). Most of these genes have also been experimentally validated as targets of FLI-1 and ETS-1 in Meg-01 (early) and CMK11-5 (late) megakaryocytic cell lines.

\( \text{PU.1} \)
PU.1 is an ETS family transcription that is essential for myeloid development and which actively inhibits erythroid development by antagonizing GATA factor function. High levels of PU.1 are responsible for the arrested development of MEL cells and its downregulation is necessary for GATA-1-mediated terminal differentiation. Recent work in zebrafish shows that tif1γ is an integral determinant of the myeloid-erythroid fate decision controlled by the gata1/pu.1 balance in the various teleost hematopoietic populations. In some hematopoietic cell types, GATA-2 is also inhibited by PU.1, however PU.1 and GATA-2 work together during the specification of mast cell fate. In megakaryocytic progenitors, PU.1 is expressed at detectable levels, although its transcription is being actively repressed by GATA-2 as GATA-2 attempts to maintain megakaryocytic fate fidelity. It is not clear what factor is driving transcription of PU.1 in this context or what role PU.1 protein plays in megakaryocyte development. PU.1 chromatin occupancy in HPC-7 cells is most similar to the binding patterns of the related ETS proteins ERG and FLI-1, although no more than ~35% of ERG or FLI-1 binding sites are co-occupied by PU.1 and PU.1 binding sites did not overlap at higher rates than would be expected by chance. Additional studies describing network interactions involving PU.1 will be instrumental in more carefully defining the role of PU.1 in hematopoietic progenitors upstream of and parallel to the classical myeloid lineages.

**MYB and microRNAs in Erythromegakaryocytic Regulation**

Clearly, the transcription factors already discussed in this review cannot account for every facet of the complex process of erythromegakaryocytic development. In particular, while transcriptional regulation can be fairly tightly controlled, post-transcriptional regulatory mechanisms, such as microRNAs, allow even finer tuning and more rapid response to stimuli. MiR-150, for example, has a role in controlling MEP fate decision that is thought to be mediated, at least in part, by negatively regulating the pro-erythroid
nuclear factor, MYB.\textsuperscript{7} Recently, another pair of microRNAs, miR-15a and -16-1, controlling MYB in human erythroid cells has been implicated in HPFH associated with trisomy 13, suggesting a role for MYB in hemoglobin switching in addition to fate decision.\textsuperscript{109} An additional microRNA, miR-126, inhibits red cell production from in vitro differentiation of human embryonic stem cells, although this report does directly test whether this effect is acting at the stage of erythromegakaryocytic fate decision.\textsuperscript{110}

**Conclusions and Future Directions**

Utilizing the extensive data obtained by many research teams, we have compiled a regulatory circuit of erythroid and megakaryocytic specification (Figure 2). In our model, expression of EKLF is indicative of an erythroid cell fate, whereas expression of FLI-1 is indicative of the megakaryocyte fate. The final cell fate decision is guided by a cumulative score of cooperative and antagonistic interactions that involve the factors discussed in this review, notably GATA-1, GATA-2, PU1, SCL, KLF1, LRF and FOG1. Interestingly, the data reviewed here make it apparent that all of the major hematopoietic factors are capable of acting as either activators or repressors, depending on the cellular context, post-translational modifications of the factor and, of course, assembled cofactors.

**Relevance and Verification of Cell Line Data**

Due to technical limitations of early ChIP-Seq methodology, most studies relied on cell lines to generate sufficient ChIP material for sequencing. In the vast majority of cases confirmed by qPCR, there is excellent correlation of binding between the cell lines and primary cells. Recently, as technology has improved, smaller numbers of cells are needed for efficient immunoprecipitation and smaller amounts of DNA are required for library preparation, allowing for comprehensive occupancy analysis in primary cells. One
recent study, which specifically examined the concordance between ChIP-Seq peaks obtained for GATA-1, SCL, and LDB1 in MEL cells with those obtained from E13.5 fetal livers, observed an overlap of 83%. A final note to consider about the use of cell lines in ChIP-Seq experiments is that for some factors, investigators are forced to use epitope tagged, overexpressed factors due to the lack of appropriate high quality antibodies.

Advances in Technology towards Future Goals

One of the ultimate (and admittedly lofty) goals of the post-genome era is a complete understanding of the code written by the billions of nucleotides that make up the human genome. This understanding is likely to manifest itself in our ability to model and specifically predict the complex molecular and biochemical interactions upon the primary structure of DNA that control gene expression. To fully understand how transcription factors select their sites and recruit activating and repressive complexes, it will be imperative that we fully understand how histone variants, histone modifications, direct DNA modifications, and even nucleosome positioning are determined, effected, and read. Recent (and presumed future) advances in technology coupled with increasing focus on miniaturization towards single cell analyses, thoughtful design of experiments to collect comprehensive datasets, and advanced computational biology techniques allow this to be an achievable, albeit daunting, task.

To gain further insight into how transcription factors select their binding sites, a thorough understanding of the determinants and competitors of nuclear factor binding must be realized. The technologies to address these basic molecular genetics questions more comprehensively and rigorously certainly exist.
First, an example of a recent implementation of high-level computational methods across multiple histone modification ChIP-Seq datasets through nine cell types, including K562 cells, identified ‘chromatin states’ defined by combinatorial patterns of histone acetylations and methylations. These data provide an additional level of information about specific genomic positions for investigators studying genomic diseases and provide a framework for understanding how histone marks work together to control chromatin structure and gene expression. Moreover, the data may be of particular interest to investigators performing ChIP-Seq for transcription factors, as it will allow them to associate peaks with specific chromatin states and infer additional biological information from their experiments.

A second discovery that may be exploited to provide new biological insights in hematopoiesis involves the 2009 rediscovery of the “sixth nucleotide,” 5-hydroxymethylcytosine (5hmC). Given the vast literature on epigenetics related to methylcytosine, it is reasonable to surmise that 5hmC will soon emerge as a regularly characterized epigenetic modification. Indeed, 5hmC imbalance has already been implicated in leukemia and several investigators have published ChIP-Seq studies detailing the genome-wide distribution of 5hmC in mouse ES cells.

Third, the use of massively parallel sequencing technology to identify nucleosome positions will likely provide new insights into how transcription factors remodel mammalian chromatin and the effects of nucleosome sliding and repositioning on transcription factor binding and site selection. Initial nucleosome positioning catalogs were constructed in yeast and nematodes due to sequencing limitations that prevented sufficient coverage for mapping in mammalian cells, but the application of these data to generate models of mammalian genomes has allowed for accurate prediction of
occupancy.119 Recent advances in sequencing hardware and software now allow for significantly more data from a single run, which implies that a complete mammalian nucleosome map will be available in the near future. In fact, a recent study of human transcription start sites reports accurate nucleosome positioning data from approximately 100 million short reads.120

Fourth, the miniaturization of the ChIP procedure itself is likely to yield important information from rare critical cell types, such as hematopoietic stem and progenitor cells. Two recent reports from the Bernstein group have approached this issue using ChIP-Seq for histone modifications in either mouse ES cells or as few as 20,000 Lin Sca "Kit" hematopoietic progenitor cells.121,122 It is likely that, as these protocols become more accessible, any cell type that can be obtained in sufficient quantities for transcript profiling will also be abundant enough for ChIP-Seq analysis.

Fifth, and finally, it is important to note that collection and publication of large discrete datasets detailing the molecular underpinnings of a single cell type at a single stage of development is not sufficient. It is imperative that the data be integrated with existing datasets to provide novel biological insights into emergent properties of the biological systems under investigation. To that end, Göttgens and colleagues have compiled a library of hematopoietic ChIP-Seq experiments. Collectively, these datasets are instructive in identifying critical hematopoietic regulatory regions, as critical factor binding sites tend to cluster and observing them en masse greatly increases one’s ability to infer biological activity from binding patterns.123 An additional example of large scale data integration comes from Novershtern et al, who recently published a study of gene expression analysis across 211 samples from 38 different prospectively isolated human primary hematopoietic cell types. With their data, they have made a webtool available
where one can perform various functions across this massive dataset. Together, these new technologies and advances in applying computational methods to biological data will allow biologists to find answers to many previously unapproachable questions about basic transcriptional regulation of cell fate and hematopoiesis.
Acknowledgements

The authors thank Aaron Dinner and S.M Ali Tabei for their assistance in generation of the regulatory circuit diagram and Tim Chlon for critically reviewing the manuscript. This review was supported by an NIH P50 award (GM081892) to the Chicago Center for Systems Biology, NIH T32-CA080621, a Malkin Family Scholar Award (L.C.D.), and a Scholar Award from the Chicago Biomedical Consortium supported by the Searle Funds at the Chicago Community Trust (L.C.D.).

Authorship

L.C.D. and J.D.C wrote the review. The authors declare that they have no competing financial interests.
REFERENCES


FIGURE LEGENDS

**Figure 1: Erythroid cells and megakaryocytes are derived from a common progenitor.** The decision of whether the MEP should give rise to red cell or platelet lineages is controlled by an array of transcription factors and microRNAs.

**Figure 2: A complex regulatory network controls MEP cell fate.** The relative levels of expression of key regulatory factors including GATA-1, GATA-2, SCL and PU1 are predicted to control the output towards erythroid cells (marked by EKLF) or megakaryocytes (marked by Fli-1). Arrows indicate gene activation and blunted lines indicate repression.
Figure 1
Figure 2
Transcription factor networks in erythroid cell and megakaryocyte development

Louis C. Doré and John D. Crispino