Published in final edited form as:

Exp Hematol. 2014 August ; 42(8): 618–629. doi:10.1016/j.exphem.2014.05.004.

Hematopoietic Transcriptional Mechanisms: From Locus-Specific to Genome-Wide Vantage Points

Andrew W. DeVilbiss^{1,2,*}, Rajendran Sanalkumar^{1,2,*}, Kirby D. Johnson^{1,2}, Sunduz Keles^{2,3}, and Emery H. Bresnick^{1,2}

¹University of Wisconsin School of Medicine and Public Health, Department of Cell and Regenerative Biology, Carbone Cancer Center, Madison, WI 53705

²UW-Madison Blood Research Program, Madison, WI 53705

³University of Wisconsin School of Medicine and Public Health, Department of Biostatistics and Medical Informatics, Madison, WI 53705

Abstract

Hematopoiesis is an exquisitely regulated process in which stem cells in the developing embryo and the adult generate progenitor cells that give rise to all blood lineages. Master regulatory transcription factors control hematopoiesis by integrating signals from the microenvironment and dynamically establishing and maintaining genetic networks. One of the most rudimentary aspects of cell type-specific transcription factor function, how they occupy a highly restricted cohort of *cis*-elements in chromatin remains poorly understood. Transformative technological advances involving the coupling of next-generation DNA sequencing technology with the chromatin immunoprecipitation assay (ChIP-seq) have enabled genome-wide mapping of factor occupancy patterns. However, formidable problems remain, notably ChIP-seq analysis yields hundreds to thousands of chromatin sites occupied by a given transcription factor, and only a fraction of the sites appear to be endowed with critical, non-redundant function. It has become en vogue to map transcription factor occupancy patterns genome-wide, while utilizing powerful statistical tools to establish correlations to inform biology and mechanisms. With the advent of revolutionary genome editing technologies, one can now reach beyond correlations to conduct definitive hypothesis testing. This review will focus on key discoveries that have emerged during the path from single loci to genome-wide analyses, specifically in the context of hematopoietic transcriptional mechanisms.

Introduction

The process of hematopoiesis involves the differentiation of hematopoietic stem cells (HSCs) into specialized progenitor cells, which give rise to the diverse blood cell types [1].

Correspondence should be addressed to: E.H.B (ehbresni@wisc.edu).

*Equal contribution

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Hematopoiesis is controlled by numerous signals and factors, which must be exquisitely coordinated to ensure the generation and maintenance of cells in all sectors of the hematopoietic hierarchy to fulfill physiological requirements and to prevent errors that underlie blood cell malignancies and non-malignant pathologies. Analogous to other complex developmental processes, cell type-specific transcription factors function through ensembles of *cis*-regulatory elements to establish and maintain genetic networks that control the genesis and maintenance of precursors, as well as lineage commitment and differentiation.

One of the complexities involved with studying hematopoiesis relates to the fact that it occurs at distinct anatomical sites during development (yolk sac, aorta gonad mesonephros (AGM) region, fetal liver, placenta and bone marrow) [2, 3]. Elucidating the underlying mechanisms necessitates careful consideration of how regulation conferred by distinct microenvironments intermeshes with cell autonomous mechanisms operational within HSCs and progenitors. Cultured cell systems that recapitulate the diverse regulatory environments of the hematopoietic niche within extraembryonic tissue and the embryo proper are not available. Thus, dissecting mechanisms underlying hematopoiesis involves considerably greater complexity than addressing comparable questions in the context of a solitary differentiated cell type. Loss-of-function studies in murine models allow one to definitively establish the requisite regulatory signals and factors, as well as the underlying mechanisms, at multiple anatomical sites in a developmentally dynamic context.

A powerful strategy to understand how cell type-specific transcription factors control biological processes involves mapping the endogenous factor occupancy sites in chromatin using chromatin immunoprecipitation (ChIP) technology [4, 5]. Combined with loss-of-function studies to define factor-regulated transcriptional profiles, this strategy reveals direct target gene ensembles. Direct target gene ensembles can be scrutinized to pinpoint potentially important physiological, cell biological, and biochemical processes regulated by the factor, and analyzing such datasets using a systems biology approach permits the assembly and modeling of regulatory networks. The advent and evolution of technologies to generate genome-wide maps of proteins bound to specific chromosomal sites has greatly catalyzed advances in our understanding of normal and malignant hematopoiesis.

Chromatin Immunoprecipitation technology and analytical strategies

The ChIP assay involves covalently crosslinking proteins to chromatin in living cells or tissues with formaldehyde, fragmenting the crosslinked chromatin, immunoprecipitating a protein of interest with an appropriate antibody, reversing the crosslinks, and using PCR to measure the recovery of specific DNA sequences, using isotype-matched antibodies as controls. The ChIP assay is not limited to analysis of proteins directly bound to DNA, as formaldehyde creates complex networks of covalently linked proteins and therefore has utility for mapping proteins indirectly associated with DNA through protein-protein interactions. Since proteins tethered to DNA via protein-protein interactions often yield considerably lower signals in the ChIP assay, bifunctional crosslinking reagents have been incorporated with formaldehyde to increase signal to noise. Similarly, in certain instances,

prolonging crosslinking time from the usual 5–10 minutes to several hours can increase signal intensity, although background signals can become unacceptable in certain instances.

The full complement of factor occupancy sites throughout a genome represents an invaluable resource that can reveal important mechanistic and biological discoveries. To achieve this goal, ChIP technology evolved to incorporate high-throughput DNA analysis methodologies. These efforts were catalyzed by the generation of microarrays harboring tiled genomic regions [6], and further complexity was achieved by incorporating large components of the genome [7]. The current state-of-the-art incorporates next-generation sequencing to identify the sequence of all immunoprecipitated DNA molecules (ChIP-seq) [8].

There are multiple robust analysis pipelines for mapping factor-chromatin interaction sites [9]. Conventional ChIP-seq data analysis involves (i) quality control analysis of sequence reads; (ii) alignment of sequence reads to a reference genome; and (iii) statistical analysis to identify regions of ChIP enrichment versus an appropriate control sample (peaks) while controlling an appropriate error rate, e.g., false discovery rate. The most notable quality control pipeline that evaluates ChIP-seq datasets in terms of overall signal to noise ratio is available from the multi-investigator genome project (ENCODE) to comprehensively map chromosome landscapes [10]. Short read alignment tools Bowtie [11] and Burrows Wheeler Alignment tool (BWA) [12], which produce robust alignments of reads that uniquely align to the genome, are most popular for step (ii).

In principle, ChIP-seq can yield a comprehensive account of factor occupancy genome-wide. In a ChIP-seq experiment, a considerable fraction (up to 30%) of the reads can align to multiple locations on the genome. Several ChIP-seq specific read alignment methods provide means of utilizing these multi-mapping reads and lead to identification of protein-DNA interaction sites in repetitive regions of genomes [13, 14]. There are a plethora of algorithms for calling peaks in ChIP-seq datasets. Some notable ones with robust performances are SPP [15], Model-based Analysis of ChIP-seq (MACS) [16], and Joint Model-based One- and Two-sample Analysis and Inference for ChIP-seq (jMOSAICS) [17]. These methods employ varying error control procedures for making lists of peaks and they often disagree on the numbers of peaks identified. A recent statistical procedure, named irreducible discovery rate (IDR) [18], bypasses this problem by deriving the number of peaks based on reproducibility in at least two replicates and stabilizes the numbers of peaks identified by different peak callers. Once peaks are identified, they are annotated with respect to their proximity to genes, and are typically divided into groups depending on whether they are distal or proximal to transcription start sites.

Another common downstream analysis of ChIP-seq data is *de novo* sequence analysis where sequence features overrepresented in peak regions are identified. This often yields consensus binding sequences for sequence-specific transcription factors. However, it can also reveal consensus sequences for factors lacking intrinsic DNA binding activity that are recruited to chromatin via protein-protein interactions. MEME-SUITE [19] provides multiple online tools for performing such analysis.

ChIP-seq studies are often designed to conform to a common template involving the following scheme: (i) delineate the genomic occupancy sites as noted above; (ii) establish the distribution of occupancy sites relative to gene structural features (e.g. promoter, intron, exon, and enhancer); (iii) identify individual and clusters of *cis*-elements enriched at occupancy sites that may mediate direct factor binding to the chromatin; and (iv) compare the genome-wide pattern with the overall chromatin binding landscape, consisting of other transcription factors, coregulators, and histone modifications. More recently, examples have emerged in which transcription factor occupancy datasets have been integrated with higher-order chromatin maps [20] generated with chromosome conformation capture technologies (e.g. 3C, Hi-C and 5C) [21, 22], although this work is still in its infancy and can involve major technical and analytical hurdles.

ChIP-seq analysis of chromatin occupancy for a transcription factor typically reveals hundreds to thousands of crosslinked (occupancy) sites, including loci that appear not to be transcriptionally regulated by the factor. As noted above, powerful analytical strategies have emerged that allow one to assemble statistically rigorous ensembles of “peaks” representing endogenous factor occupancy sites. While this technology is undoubtedly sophisticated, the interpretation of ChIP-seq datasets remains highly challenging. Formidable challenges include: (i) how to sift through abundant peaks to identify functionally critical sites; (ii) how to establish direct target genes with high fidelity; and (iii) how to discriminate between *cis*-elements mediating factor occupancy and function through chromatin versus nonfunctional *cis*-elements residing at or near the occupancy site. As *cis*-elements are often less than 10 bp, they are extremely abundant within a genome, and certain elements can be rendered nonfunctional based on the local and higher-order chromatin environment. Nearly every chromosomal region contains sequences that may be mis-identified as functional *cis*-elements. It is reasonable to assume that some of these *cis*-elements represent crucial, non-redundant transcriptional regulatory *cis*-elements, while others may not be critical and/or may function redundantly with other elements. Elucidating mechanisms that endow specific DNA sequences with non-redundant transcriptional regulatory activity at endogenous loci represents an additional major challenge. Addressing these issues will ultimately yield an integrative perspective on how stem and progenitor cells give rise to the hematopoietic system, and obviously the further development of the respective technologies will impact favorably upon a wide spectrum of fields. However, transformative advances will likely be required to achieve this lofty goal.

From chromatin occupancy to mechanistic and biological insights

Prior to the era of genome-wide chromatin analysis, in-depth studies on individual loci, including the erythroid cell-specific β -globin locus, revealed important mechanistic insights and provided a rigorous intellectual framework to guide genome-wide studies. Dissecting mechanisms underlying erythroid gene expression led to the purification and cloning of the cell type-specific transcription factor GATA-1, based on its capacity to bind and regulate β -globin *cis*-elements and *cis*-elements associated with a broad ensemble of erythroid genes [23–26]. Targeted deletion of mouse *Gata1* revealed its essential function to control erythropoiesis [27–29] and also importance for regulating megakaryopoiesis [30], as well as eosinophil [31] and mast cell biology [32]. The discovery of GATA-1 spawned the cloning

of a new family of dual zinc finger transcription factors (GATA proteins) consisting of six mammalian proteins that control the development and function of critical processes including blood, heart, brain, vasculature, and adipose, as well as playing a role in human diseases, such as anemia, breast cancer, and leukemia [33–38].

Analyzing the histone modification state of the β -globin chromatin domain in erythroid and non-erythroid cells revealed a specific pattern of modifications that selectively demarcated functionally unique gene regions [39, 40]. Certain factors, e.g. GATA-1 and the basic-leucine zipper transcription factor NF-E2, are determinants of distinct components of the pattern [41–43], illustrating the importance of combinatorial interactions between multiple transcription factors to establish and/or maintain histone modification patterns within chromatin domains. Analyzing proteins functioning through the β -globin chromatin insulator, which confers enhancer blocking and “barrier” activities, led to the discovery of a previously studied transcriptional repressor, CTCF, as a functionally important insulator binding component [44]. This work spawned genome-wide studies demonstrating a broad role of CTCF in organizing the three-dimensional folding of the genome [45]. Evaluating how the distal β -globin locus control region, a collection of four DNaseI hypersensitive sites containing many *cis*-elements [46, 47], communicates with downstream promoters provided evidence for dynamically regulated chromatin looping [48] and gene positioning [49] within the three-dimensional constraints of the cell nucleus. Studies on the regulation of the *Gata2* gene led to the discovery that multiple GATA-1/2-bound *cis*-elements function in qualitatively distinct modes at this endogenous locus, despite sharing fundamental structural properties [50–53] (discussed below in detail). These examples illustrate significant mechanistic advances that preceded the advent and implementation of genome-wide technologies and established a conceptual framework for devising genome-wide strategies and interpreting the resulting datasets.

GATA-1 was among the earliest of mammalian transcription factors to be analyzed by genomic technologies [54–57]. Although GATA-1 was discovered in erythroid cells, it is also expressed in a select number of additional cell types including megakaryocytes, mast, and eosinophil cells. GATA-1 preferentially binds WGATAR motifs in naked DNA [58, 59]. Although this sequence occurs approximately 7 million times throughout the mouse genome [60], GATA-1 occupies a highly restricted set (0.1 – 1%) of these elements in cultured and primary erythroid cells [55, 61]. Analysis of *cis*-elements at GATA-1 occupancy sites genome-wide yielded the refined consensus (C/G)(A/T)GATAA(G/A/C)(G/A/C), and the percentage of these sequences occupied in the genome increases several fold to 0.7% versus occupancy at WGATAR motifs [55]. Thus, GATA motifs do not suffice for chromatin occupancy, at least based on state-of-the-art technology for measuring chromatin occupancy. This lack of sufficiency suggests that other factors, including interacting proteins, local chromatin landscape, higher-order chromatin structure, and subnuclear positioning, are also determinants of the highly constrained occupancy pattern.

GATA-1 establishes the erythroid cell transcriptome through both transcriptional activation and repression of target genes [62, 63]. Mechanisms governing the decision of whether GATA-1 activates or represses a target gene have been studied using various approaches, including ChIP-seq technology [55–57]. GATA-1 utilizes the nine zinc finger coregulator

protein Friend of GATA-1 (FOG-1) to activate and repress the majority of its target genes [63, 64]. Though some GATA-1 target genes are FOG-1-insensitive [63, 65, 66], the biological importance of FOG-1-independent GATA-1 function is not known. FOG-1 is recruited to chromatin via direct interaction with GATA-1 and lacks intrinsic DNA binding activity [63]. FOG-1 associates with the repressive Nucleosome Remodeling and Deacetylase (NuRD) complex, which contains the ATPase CHD4 (Mi2 β) [67]. Mi2 β occupies GATA-1-bound chromatin sites at GATA-1-repressed and -activated loci [68]. The histone acetyltransferases CREB-binding protein and its paralog p300 (CBP/p300) bind GATA-1 and facilitate GATA-1-mediated activation of target genes [69, 70]. CBP/p300 also occupy GATA-1-activated and -repressed genes [71, 72]. The location and frequency of GATA-1 occupied sites correlate with the transcriptional output of the target gene. GATA-1-induced genes are more likely to exhibit GATA-1-occupancy close to the transcription start site, whereas GATA-1 occupies distal sites more frequently at repressed genes. In addition, multiple GATA-1-occupied sites at a locus correlates with transcriptional activation [57]. One needs to be cautious about overinterpreting global correlations, however, as there are many examples of functionally critical, GATA-1-activated genes with distal GATA-1 occupancy sites. It is essential to develop definitive mechanistic insights, rather than inferring function from correlations. In this regard, it is instructive to consider the cohort of transcription factors and coregulators implicated in mediating GATA-1 functions.

GATA-1 co-occupies chromatin with several transcription factors and coregulators, and multi-factor occupancy may have utility for predicting the functional importance of occupied *cis*-elements and/or unique functions. In erythroid and megakaryocytic lineages, Scl/TAL1 binding sites are non-randomly enriched at GATA-1-occupied sites. The basic-helix-loop-helix protein Scl/TAL1 positively regulates HSC function [73, 74] and also the development and function of erythroid cells [75]. A heterodimer of Scl/TAL1 and E2A proteins binds E-boxes with the consensus sequence CANNTG [76, 77]. Wadman et al. [78] described a composite element with the configuration CANNTG - 8 bp - AGATAA that binds GATA-1, Scl/TAL1 and additional non-DNA binding factors. We and others have observed GATA factor-occupied composite elements with this configuration and additional configurations with a variable GATA motif and spacer length. While direct DNA binding is assumed to be fundamental for Scl/TAL1 function, Scl/TAL1 can be recruited to chromatin by other factors, including GATA-1, and certain Scl/TAL1 activities *in vivo* are unaffected by a DNA binding-crippling mutation [79]. Scl/TAL1 occupancy at GATA-1-bound sites is associated with active transcription, whereas a lack or loss of Scl/TAL1 occupancy correlates with GATA-1-mediated repression [56, 57, 72].

Scl/TAL1-E2A and GATA-1 form a multimeric complex on DNA with LDB1 [78], a mammalian ortholog of Chip [80], which mediates long-range genetic interactions, and LMO2, a leukemogenic LIM domain protein [81]. Conditional deletion of *Ldb1* in the hematopoietic systems depletes fetal and adult HSCs, disrupts definitive erythropoiesis and megakaryopoiesis, and depletes hemangioblasts, which give rise to endothelium and primitive erythroid cells [82–85]. Consistent with the multimeric complex model noted above, GATA-1, Scl/TAL1 and LDB1 commonly co-occupy chromatin, although additional sites

appear to lack GATA-1. GATA-1, Scl/TAL1, and LDB1 co-occupancy correlates with transcriptional activation [57]. Similar to GATA-1, LDB1 frequently occupies intronic chromatin sites. LDB1-occupied sites harbor motifs for other transcription factors with important erythroid functions, including Runx1, Foxo3, Egr1, Nfe2, Sp1, and Ets1 [86]. LDB1 directly regulates many erythroid-specific genes, including genes encoding transcription factors, cytoskeletal components, globin chains, heme biosynthesis enzymes, and erythroid surface marker genes [84].

The erythroid-specific transcription factor KLF1 (originally termed EKLF) binds CCNCNCCCN motifs, and establishes the erythroid transcriptome in cooperation with GATA-1, and possibly independently [87–93]. *Klf1*^{-/-} erythroid cells fail to undergo terminal maturation during fetal liver erythropoiesis and exhibit defects in chromatin modification and cell cycle progression [93, 94]. GATA-1-binding sites are non-randomly enriched near KLF1-binding motifs, most commonly with a displacement of 18 nucleotides [91]. Two different KLF1 genome-wide occupancy studies yielded conflicting results as to the extent of GATA-1 and KLF1 co-occupancy [88, 91]. This discrepancy is likely explained, in part, by vastly different numbers of statistically significant KLF1 occupancy sites analyzed in the two studies. The extent of co-localization between transcription factors on chromatin derived from ChIP-seq data needs to be carefully interpreted, as differences in sequence read numbers and sequence read quality can greatly influence peak identification. KLF1 occupancy undergoes dynamic changes during erythroid differentiation [88]. In erythroid progenitors, the majority of KLF1 is associated with promoters of active genes. However, as progenitor cells differentiate into erythroblasts, KLF1 chromatin occupancy shifts to coding regions and intergenic regions and is most commonly associated with repressed genes. KLF1 direct targets include genes critical for erythrocyte function, such as the globin genes, heme biosynthetic enzymes, cytoskeletal components, and cell membrane components, many of these genes overlapping with direct GATA-1 target genes [88, 90, 91]. GATA-1-Scl/TAL1 co-occupied sites and GATA-KLF1 co-occupied sites rarely coincide [84, 91], and KLF1 and LDB1 co-occupied sites almost never overlap with GATA-1-occupied sites lacking Scl/TAL1 [84]. These genomic analyses of multi-factor occupancy revealed that KLF1 establishes components of the erythroid cell transcriptome that overlap with, and may also be distinct from, that of GATA-1.

The related GATA factor, GATA-2, is expressed in HSCs, hematopoietic progenitors, endothelial cells, and neurons. GATA-2 is a crucial determinant of HSC genesis and function [52, 95–98]. Erythrocyte differentiation involves the progressive maturation of immature erythroblasts into mature erythroblasts, which undergo enucleation to form reticulocytes and then erythrocytes [99]. Immature erythroblasts express GATA-2, and as GATA-1 is upregulated during erythroid maturation, GATA-1 utilizes FOG-1 to directly repress *Gata2* transcription [63, 100, 101]. *Gata2* repression involves GATA-1-mediated displacement of GATA-2 from *Gata2* locus sites, a process termed the GATA switch [35, 36].

At the genomic level, GATA-2 can occupy certain chromatin sites in hematopoietic precursors cells, which are subsequently occupied by GATA-1 during erythroid maturation [55, 102, 103]. GATA switches occur at many sites throughout the genome and are

frequently associated with altered transcriptional output – either repression or activation [35, 36]. In the G1ME megakaryocytic cell line, GATA switching is widespread and associated with transcriptional activation and repression [104]. By contrast, in multipotent FDCPmix cells, which have some capacity to differentiate into erythroid cells, GATA switching is less frequent, and in certain contexts, GATA-1 and GATA-2 apparently co-occupy the same chromatin sites [105]. This incomplete switching may reflect the limitation of a factor that promotes GATA switches, e.g. the erythroid GATA-1 co-regulator FOG-1 [106], and/or other limitations of these cells, which do not efficiently undergo erythroid maturation.

Transitioning from chromatin occupancy to function at endogenous loci

Though GATA switch sites contain conserved GATA motifs and additional conserved *cis*-elements, are occupied by GATA factors and other factors *in vivo*, and contain chromatin attributes commonly believed to demarcate transcriptional enhancers, targeted deletion studies have revealed major unexpected qualitative and quantitative differences in GATA switch site function [50–53, 107]. These *cis*-element knockout studies have been particularly instructive at the *Gata2* locus (Fig. 1). Using homologous recombination in murine ES cells, mutant mouse strains lacking the five GATA switch sites individually were generated [50–53, 107]. *In vivo* analyses were conducted to determine if the deletions induced embryonic lethality, altered GATA-2 expression in primary hematopoietic cells and tissues, deregulated genetic networks known to regulate hematopoiesis, and changed HSC, progenitor and/or differentiated blood cell levels.

Deletion of the –2.8 site (2800 bp upstream of the *Gata2* 1S promoter), which contains multiple GATA motifs, caused a small decrease in *Gata2* expression in select hematopoietic precursors [50]. However, the deletion did not significantly impact on hematopoiesis. Deletion of the GATA motif palindrome-containing –1.8 site resulted in *Gata2* derepression in late-stage erythroblasts [53]. However, embryonic and adult hematopoiesis were normal in the mutant. Despite multiple conserved binding motifs, factor occupancy in primary cells, and enhancer attributes, the –3.9 site deletion did not influence *Gata2* expression or hematopoiesis [107]. By contrast, deletion of the +9.5 site, which harbors an intronic E-box-spacer-GATA composite element, yielded embryonic lethality, inability of hemogenic endothelium in the AGM to generate HSCs and ablation of the fetal liver hematopoietic stem/progenitor cell compartment, and a vascular integrity defect [51, 52]. Mechanistically, the +9.5 site deletion reduced *Gata2* expression in the AGM, which was linked to disruption of a gene network containing established regulators of hemogenic endothelium and HSCs. As the vast majority of genomic E-box-spacer-GATA composite elements are not occupied by GATA-1 (or other GATA factors), the mere presence of a composite element is insufficient to generate a critical GATA binding element. It will be instructive to deploy genome editing to interrogate a broad spectrum of GATA switch sites at distinct loci.

While knockout of the –77 GATA switch site (containing multiple GATA motifs) in mice has not been described, this region of the human *GATA2* locus is rearranged in an aggressive form of acute myeloid leukemia with poor prognosis [Inv(3)(q21q26.2)] [108]. The rearrangement positions the –77 site in close proximity to the gene encoding the oncogenic transcription factor *Evi1*, resulting in *Evi1* upregulation and leukemogenesis [108, 109].

Whether or not the -77 is an essential mediator of endogenous *Gata2* regulation and hematopoiesis, this scenario exemplifies how a *cis*-element can be coopted by an oncogene as a mechanism underlying human leukemogenesis. Importantly, while ChIP-seq approaches in certain systems detect GATA factor occupancy at the $+9.5$ or -77 GATA switch sites, the occupancy is not distinctive relative to that at the -1.8 , -2.8 , and -3.9 sites, which lack essential nonredundant functions.

Analogous to the *Gata2* work, rigorous studies were conducted to dissect the importance of the individual DNaseI hypersensitive sites (HS1–HS4) of the murine β -globin locus control region (LCR) [110–115]. Studies in transgenic mice indicated that the LCR confers copy number-dependent and position-independent expression of linked transgenes integrated at ectopic loci [46, 47]. Targeted deletions of the individual DNaseI hypersensitive sites at the endogenous locus [110–115] revealed results that were not predictable from transfection and transgenic mouse assays. Despite compelling enhancer activities in these assays, certain qualitative differences in their properties, and extensive factor occupancy, deleting the individual hypersensitive sites of the LCR only modestly reduced β -like globin transcription and did not reveal obvious qualitative differences. While our goal is not to survey all *cis*-element knockouts, an important conclusion from the type of work conducted at the *Gata2* and *β -globin* loci is that genetic analysis of occupied *cis*-elements at endogenous loci is essential to assess physiological function.

Discovering functional *cis*-elements occupied by hematopoietic transcription factor complexes

Since ascertaining the functional importance of factor-occupied chromatin sites detected by ChIP is particularly challenging, new approaches are required to fast-forward the winnowing and sifting through hundreds or thousands of occupied sequences. The advent of facile genome editing technologies, namely Transcription Activator-like Effector Nucleases (TALENs) [116, 117] and the CRISPR/Cas9 system [118], represent transformative developments that allow one to excise and/or mutate *cis*-elements at endogenous loci as a direct assessment of *cis*-element function in a physiological context. Though these technologies permit definitive functional tests, given the large number of occupied sites detected in genome-wide analyses, effective prioritization strategies remain critical.

A logical approach to prioritize ChIP-seq data, with the goal of identifying functionally critical target genes and *cis*-elements, involves identifying ensembles of factors that co-occupy a common chromatin site containing evolutionarily conserved *cis*-elements. The inherent assumption is that multi-factor occupancy at a site has a higher probability of translating into functional importance versus occupancy involving a single factor or a more restricted number of factors [83, 119–121]. A “heptad” of transcription factors [Scl/TAL1, GATA-2, RUNX1, ERG (Ets factor), LYL1 (basic-helix-loop-helix factor), LMO2, and FLI1 (Ets factor)] co-occupy differentially expressed genes in the multipotent HPC7 hematopoietic cell line [120], and this result was extended to primary human HSPCs [121]. Furthermore, GATA-1, Scl/TAL1, GATA-2, FLI1 and RUNX1 co-occupy megakaryocyte genes [119].

Multi-factor co-occupancy can facilitate the process of pinpointing functionally important genes and *cis*-elements, and provides a foundation for investigating the dynamics of combinatorial occupancy during HSC generation, maintenance, lineage specification, and maturation. Almost nothing is known about the order-of-assembly and disassembly of complexes and how extracellular cues from the hematopoietic niche trigger or block these mechanistic steps. Data mining or experimental analysis can be used to survey for factors not known to associate with a particular complex. Such an association may imply functional importance, and therefore provides justification for further biological and/or mechanistic studies.

ETS (E26 transformation specific) proteins constitute a large class of transcription factors [122] that frequently co-localize with hematopoietic transcription factors. ETS factors bind simple GGA(A/T) motifs on naked DNA, which are enriched at GATA-2 occupancy sites in endothelial [123] and megakaryocytic [104] cells. In addition, a human immunodeficiency disorder termed MonoMAC Syndrome, which was initially demonstrated to be caused by mutations of the GATA-2 DNA binding zinc finger [124–127], can result from disruption of the *Gata2* +9.5 composite element [51] or point mutations within a neighboring Ets site [128]. As this ETS site is required for +9.5 site enhancer activity in a transient transfection assay, presumably the composite element-bound GATA-Scf/TAL1 complex requires a neighboring ETS factor to generate a functional complex. The ETS factor PU.1 (*Sfp1*) is essential for generating T and B lymphocyte progenitors, monocytes, and granulocytes, and facilitates terminal maturation of fetal liver erythrocytes [129, 130]. PU.1 interacts with lineage-restricted co-factors to instigate B cell or macrophage lineage commitment programs [131]. PU.1 chromatin occupancy differs in macrophages and B cells, although certain occupancy sites are shared in both cell types [132]. The shared peaks are enriched at proximal promoter regions, and cell-type specific peaks are more frequent at distal regions. The differentially occupied chromatin sites in B cells and macrophages harbor unique enrichments of binding motifs for E2a, EBF and AP-1, and C/EBP. These results suggest that PU.1 establishes the transcriptomes of different cell types via differential enhancer occupancy and/or function. In erythroid precursors, PU.1 opposes GATA-1 activity and therefore suppresses erythroid maturation in certain contexts [133, 134].

Other ETS factors have emerged with crucial roles to control hematopoiesis, including ER71(ETV2), which is expressed in hemangioblasts, precursors to endothelial and hematopoietic cells. Loss-of-function studies indicate an essential role for ER71(ETV2) to specify hematovascular precursors. It will be particularly instructive to establish the direct target gene ensembles and the corresponding genetic networks [135–138].

Cis-elements that bind the Activator Protein-1 (AP-1) family of transcription factors [139] are highly enriched near GATA-2 binding sites in endothelial but not erythroid cells [123]. GATA-2 is necessary for maximal expression of endothelial-specific AP-1 target genes [123, 140]. As AP-1 is directly targeted by diverse cell signaling mechanisms, it is attractive to consider how extracellular signals targeting AP-1 might control GATA factor multimeric complex assembly and/or function. Studies are in their infancy to elucidate mechanisms underlying multi-factor chromatin occupancy and function. Elucidating all relevant proteins,

their interactions, and the mechanistic basis for their activities continues to be of high priority.

Individual master regulatory transcription factors can orchestrate cell type-specific factor occupancy genome-wide, suggesting that different components of a multimeric complex on chromatin might be differentially important for promoting complex assembly. In an Embryonic stem (ES) cell differentiation system that models hemogenic endothelial cell function, RUNX1 controls Scl/TAL1 and FLI1 occupancy patterns [141]. In certain instances, multimeric factor complexes may reside on chromatin sites at target loci prior to the physiological demand to regulate the respective gene. The expression of lineage-specific factors in HSCs and multipotent stem cells is believed to represent a “priming” step to ensure efficient differentiation [142–144]. Certain transcription factors are termed “pioneer factors”, as they have the capacity to occupy DNA motifs on the surface of a nucleosome [145]. Once bound, the pioneer factor would recruit coregulators that stimulate local chromatin modification/remodeling, which in turn would provide the requisite accessibility for other factors to bind. In the GATA factor switching paradigm during erythroid differentiation [35, 36, 146], it is attractive to propose that GATA-2 may prime the switch sites, thus creating favorable sites for subsequent GATA-1 binding and altered transcriptional output. In addition to priming *cis*-regulatory elements for subsequent lineage-specific expression, lineage-restricted factors can maintain lineage identity by “bookmarking” the occupied chromatin sites during the cell cycle, thus maintaining a state poised for subsequent transcriptional activity [147].

Lessons learned from genome-wide chromatin occupancy analyses

An enormous barrage of genome-wide chromatin occupancy data for hematopoietic transcription factors, chromatin modifying and remodeling enzymes, and histone modifications has been accrued by the ENCODE project [10, 148, 149] and individual research groups. The ENCODE project has provided easy access to valuable datasets. By virtue of the goals of the ENCODE project to collect and archive genomic data without drilling deeply into mechanisms, one needs to carefully scrutinize and validate the data. A compendium of validated ChIP-seq data from mouse [150, 151] and human [152] samples contains datasets for diverse hematopoietic transcription factors in primary cells and cell lines. By comparing datasets generated from cell lines and the respective primary cells, strong evidence has emerged for the utility of relevant cell lines as models for specific aspects of hematopoiesis; factor occupancy profiles and gene expression profiles can be strikingly similar [55, 61, 105]. However, since hematopoiesis occurs at distinct anatomical sites with unique microenvironments, many questions require *in vivo* analysis.

While ChIP-seq analysis in primary cells has obvious advantages, significant challenges remain. Cell isolation techniques often do not yield highly pure populations, and generating sufficient material for genomic analysis when dealing with rare cell types such as HSCs is difficult and expensive. Though most ChIP-seq protocols require large numbers of cells to ensure quality sequence reads, recent advances have led to improved genome-wide profiling methods that utilize considerably lower cell numbers. A typical ChIP-seq protocol requires millions of cells per condition. In principle, technologies like single-tube linear DNA

amplification (LinDA) ([153], nanoChIP-seq [154] or native ChIP-seq methods [155] reduce the input cell requirement to thousands. Alternatively, ChIP-seq analyses can be conducted in surrogate systems – either a distinct type of primary cell or biologically relevant cell line. For example, gene expression analysis in single HSCs (as defined by cell surface marker expression) revealed that *Gata2* and *Gfi1* expression negatively correlate. ChIP analysis conducted in primary mast cells revealed Gfi1 occupancy of a *cis*-element –83 kb upstream of the *Gata2* promoter, and Gfi1 represses the reporter activity of a –83 kb site–*LacZ* construct in HPC7 cells. Mast cells were chosen for this ChIP-seq analysis because they express the stem cell factor receptor (*c-Kit*) and certain transcription factors that also function in HSCs [156]. While the –83 kb site has not been demonstrated to function as an enhancer in HSCs via genetic deletion, these results demonstrate that relevant surrogate cell types can be employed to identify important mechanisms in cells that are too scarce to conduct genome wide occupancy profiling. Further improvements in technologies to generate genome-wide factor occupancy profiles from small numbers of cells will continue to refine and extend knowledge on the chromatin landscape *in vivo*.

A major goal of ChIP-seq studies is to establish the direct target gene ensemble for a factor or collection of factors. Traditionally, ChIP-seq peaks are assigned to genes based on proximity to a transcription start site or gene body. Merging chromatin occupancy data with loss-of-function and/or genetic complementation data that establishes factor-regulated gene expression strengthens the direct target gene annotation [55, 157]. Limitations relate to the fact that *cis*-elements can reside tens of thousands of kilobases from genes [158] and perhaps even on different chromosomes [159]. Certain *cis*-elements function over long distances on chromosomes via chromatin looping [160]. Methodologies like chromatin confirmation capture (3C) and high-throughput variations (5C, Hi-C) [161] were developed to identify long-range chromatin interactions. Merging ChIP and 3C technologies (ChIP-loop) [162] and its genomic version Chromatin Interaction Analysis by Paired-End Sequencing (ChIA-PET) [163, 164] can further facilitate target gene identification. More broadly, genome-wide occupancy data can be merged with histone profiles, chromatin accessibility patterns, and/or other enhancer attributes to predict potential active enhancers regulated by the factor of interest. However, certain *cis*-elements bearing known hallmarks of enhancers, e.g. the *Gata2* –3.9 GATA switch site, can be deleted from the genome with no measurable molecular, cellular, or physiological phenotypes [107].

ChIP technology has been revolutionary in the sense that it provides a facile approach to visualize protein interactions at specific sites within complex genomes. *In vivo* footprinting methodologies provided evidence for alterations in accessibility of restricted chromatin sites, which were interpreted with respect to possible factor occupancy. However, these methods were difficult and often culminated in inferences about the potential factor bound. While multi-factor occupancy and predictions based on criteria such as evolutionary conservation, chromatin status, and other molecular attributes represent instructive parameters to gauge *cis*-element function, advances in gene editing now allow one to directly assess function and therefore to move beyond correlations. Thus, the application of this transformative technology to address innovative questions will almost certainly generate paradigm-shifting

discoveries that shape the future of basic and translational science vis-à-vis hematopoietic transcriptional mechanisms.

Acknowledgments

This work was supported in part by NIH grants R37DK50107 and R01DK68634 and support from the Midwest Athletes Against Childhood Cancer. A.W.D. is a predoctoral fellow of the American Heart Association.

References

1. Ema H, Morita Y, Suda T. Heterogeneity and hierarchy of hematopoietic stem cells. *Exp Hematol*. 2014; 42:74–82. e72. [PubMed: 24269919]
2. Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat Immunol*. 2008; 9:129–136. [PubMed: 18204427]
3. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*. 2008; 132:631–644. [PubMed: 18295580]
4. Dedon PC, Soultis JA, Allis CD, Gorovsky MA. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. *Anal Biochem*. 1991; 197:83–90. [PubMed: 1952079]
5. Im H, Grass JA, Johnson KD, Boyer ME, Wu J, Bresnick EH. Measurement of protein-DNA interactions in vivo by chromatin immunoprecipitation. *Methods Mol Biol*. 2004; 284:129–146. [PubMed: 15173613]
6. Sun LV, Chen L, Greil F, et al. Protein-DNA interaction mapping using genomic tiling path microarrays in *Drosophila*. *Proc Natl Acad Sci U S A*. 2003; 100:9428–9433. [PubMed: 12876199]
7. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev*. 2002; 16:235–244. [PubMed: 11799066]
8. Robertson G, Hirst M, Bainbridge M, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods*. 2007; 4:651–657. [PubMed: 17558387]
9. Bailey T, Krajewski P, Ladunga I, et al. Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS computational biology*. 2013; 9:e1003326. [PubMed: 24244136]
10. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012; 22:1813–1831. [PubMed: 22955991]
11. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
13. Chung D, Kuan PF, Li B, et al. Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. *PLoS computational biology*. 2011; 7:e1002111. [PubMed: 21779159]
14. Newkirk D, Biesinger J, Chon A, Yokomori K, Xie X. AREM: aligning short reads from ChIP-sequencing by expectation maximization. *Journal of computational biology: a journal of computational molecular cell biology*. 2011; 18:1495–1505. [PubMed: 22035330]
15. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology*. 2008; 26:1351–1359.
16. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9:R137. [PubMed: 18798982]
17. Zeng X, Sanalkumar R, Bresnick EH, Li H, Chang Q, Keles S. jMOSAiCS: joint analysis of multiple ChIP-seq datasets. *Genome Biol*. 2013; 14:R38. [PubMed: 23844871]
18. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*. 2011; 5:1752–1779.

19. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37:W202–208. [PubMed: 19458158]
20. Lan X, Witt H, Katsumura K, et al. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* 2012; 40:7690–7704. [PubMed: 22675074]
21. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science.* 2002; 295:1306–1311. [PubMed: 11847345]
22. Dekker J. The three ‘C’s of chromosome conformation capture: controls, controls, controls. *Nat Methods.* 2006; 3:17–21. [PubMed: 16369547]
23. Evans T, Reitman M, Felsenfeld G. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci U S A.* 1988; 85:5976–5980. [PubMed: 3413070]
24. Evans T, Felsenfeld G. The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell.* 1989; 58:877–885. [PubMed: 2776214]
25. Tsai SF, Martin DI, Zon LI, D’Andrea AD, Wong GG, Orkin SH. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature.* 1989; 339:446–451. [PubMed: 2725678]
26. Yamamoto M, Ko LJ, Leonard MW, Beug H, Orkin SH, Engel JD. Activity and tissue-specific expression of the transcription factor NF- E1 multigene family. *Genes Dev.* 1990; 4:1650–1662. [PubMed: 2249770]
27. Pevny L, Simon MC, Robertson E, et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature.* 1991; 349:257–260. [PubMed: 1987478]
28. Simon MC, Pevny L, Wiles MV, Keller G, Costantini F, Orkin SH. Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. *Nat Genet.* 1992; 1:92–98. [PubMed: 1302015]
29. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A.* 1996; 93:12355–12358. [PubMed: 8901585]
30. Shivdasani RA, Fujiwara Y, McDevitt MA, Orkin SH. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J.* 1997; 16:3965–3973. [PubMed: 9233806]
31. Yu C, Cantor AB, Yang H, et al. Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J Exp Med.* 2002; 195:1387–1395. [PubMed: 12045237]
32. Miglaccio AR, Rana RA, Sanchez M, et al. GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant. *J Exp Med.* 2003; 197:281–296. [PubMed: 12566412]
33. Charron F, Nemer M. GATA transcription factors and cardiac development. *Semin Cell Dev Biol.* 1999; 10:85–91. [PubMed: 10355032]
34. Orkin SH. GATA-binding transcription factors in hematopoietic cells. *Blood.* 1992; 80:575–581. [PubMed: 1638017]
35. Bresnick EH, Katsumura KR, Lee HY, Johnson KD, Perkins AS. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res.* 2012; 40:5819–5831. [PubMed: 22492510]
36. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, Keles S. GATA switches as developmental drivers. *J Biol Chem.* 2010; 285:31087–31093. [PubMed: 20670937]
37. Kumar MS, Hancock DC, Molina-Arcas M, et al. The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell.* 2012; 149:642–655. [PubMed: 22541434]
38. Kouros-Mehr H, Bechis SK, Slorach EM, et al. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell.* 2008; 13:141–152. [PubMed: 18242514]

39. Forsberg EC, Downs KM, Christensen HM, Im H, Nuzzi PA, Bresnick EH. Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc Natl Acad Sci U S A*. 2000; 97:14494–14499. [PubMed: 11121052]
40. Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J*. 1994; 13:1823–1830. [PubMed: 8168481]
41. Johnson KD, Grass JA, Boyer ME, et al. Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. *Proc Natl Acad Sci U S A*. 2002; 99:11760–11765. [PubMed: 12193659]
42. Kiekhäfer CM, Grass JA, Johnson KD, Boyer ME, Bresnick EH. Hematopoietic-specific activators establish an overlapping pattern of histone acetylation and methylation within a mammalian chromatin domain. *Proc Natl Acad Sci U S A*. 2002; 99:14309–14314. [PubMed: 12379744]
43. Kiekhäfer CM, Boyer ME, Johnson KD, Bresnick EH. A WW domain-binding motif within the activation domain of the hematopoietic transcription factor NF-E2 is essential for establishment of a tissue-specific histone modification pattern. *J Biol Chem*. 2004; 279:7456–7461. [PubMed: 14597626]
44. Chung JH, Whiteley M, Felsenfeld G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell*. 1993; 74:505–514. [PubMed: 8348617]
45. Ohlsson R, Bartkuhn M, Renkawitz R. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma*. 2010; 119:351–360. [PubMed: 20174815]
46. Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res*. 1987; 15:10159–10177. [PubMed: 3480506]
47. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent high-level expression of the human b-globin gene in transgenic mice. *Cell*. 1987; 51:975–985. [PubMed: 3690667]
48. Vakoc CR, Letting DL, Gheldof N, et al. Proximity among distant regulatory elements at the beta globin locus requires GATA-1 and FOG-1. *Mol Cell*. 2005; 17:453–462. [PubMed: 15694345]
49. Ragozy T, Telling A, Sawado T, Groudine M, Kosak ST. A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. 2003; 11:513–525.
50. Snow JW, Trowbridge JJ, Johnson KD, et al. Context-dependent function of “GATA switch” sites in vivo. *Blood*. 2011; 117:4769–4772. [PubMed: 21398579]
51. Johnson KD, Hsu AP, Ryu MJ, et al. Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J Clin Invest*. 2012; 122:3692–3704. [PubMed: 22996659]
52. Gao X, Johnson KD, Chang YI, et al. Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J Exp Med*. 2013; 210:2833–2842. [PubMed: 24297994]
53. Snow JW, Trowbridge JJ, Fujiwara T, et al. A single cis element maintains repression of the key developmental regulator Gata2. *PLoS genetics*. 2010; 6:e1001103. [PubMed: 20838598]
54. Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M. GATA-1 binding sites mapped in the beta-globin locus using mammalian ChIP-chip analysis. *Proc Natl Acad Sci U S A*. 2002; 99:2924–2929. [PubMed: 11867748]
55. Fujiwara T, O'Geen H, Keles S, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell*. 2009; 36:667–681. [PubMed: 19941826]
56. Yu M, Riva L, Xie H, et al. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell*. 2009; 36:682–695. [PubMed: 19941827]

57. Cheng Y, Wu W, Kumar SA, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res.* 2009; 19:2172–2184. [PubMed: 19887574]
58. Ko LJ, Engel JD. DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol.* 1993; 13:4011–4022. [PubMed: 8321208]
59. Merika M, Orkin SH. DNA-binding specificity of GATA family transcription factors. *Mol Cell Biol.* 1993; 13:3999–4010. [PubMed: 8321207]
60. Katsumura KR, DeVilbiss AW, Pope NJ, Johnson KD, Bresnick EH. Transcriptional mechanisms underlying hemoglobin synthesis. *Cold Spring Harbor perspectives in medicine.* 2013; 3:a015412. [PubMed: 23838521]
61. Kang YA, Sanalkumar R, O'Geen H, et al. Autophagy driven by a master regulator of hematopoiesis. *Mol Cell Biol.* 2012; 32:226–239. [PubMed: 22025678]
62. Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood.* 2004; 104:3136–3147. [PubMed: 15297311]
63. Crispino JD, Lodish MB, MacKay JP, Orkin SH. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell.* 1999; 3:219–228. [PubMed: 10078204]
64. Tsang AP, Visvader JE, Turner CA, et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell.* 1997; 90:109–119. [PubMed: 9230307]
65. Johnson KD, Boyer ME, Kang JA, Wickrema A, Cantor AB, Bresnick EH. Friend of GATA-1-independent transcriptional repression: a novel mode of GATA-1 function. *Blood.* 2007; 109:5230–5233. [PubMed: 17339418]
66. DeVilbiss AW, Boyer ME, Bresnick EH. Establishing a hematopoietic genetic network through locus-specific integration of chromatin regulators. *Proc Natl Acad Sci U S A.* 2013; 110:E3398–3407. [PubMed: 23959865]
67. Hong W, Nakazawa M, Chen YY, et al. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J.* 2005; 24:67–78.
68. Miccio A, Wang Y, Hong W, et al. NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. *EMBO J.* 2010; 29:442–456. [PubMed: 19927129]
69. Blobel GA, Nakajima T, Eckner R, Montminy M, Orkin SH. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci USA.* 1998; 95:2061–2066. [PubMed: 9482838]
70. Hung HL, Lau J, Kim AY, Weiss MJ, Blobel GA. CREB-binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Mol Cell Biol.* 1999; 19:3496–3505. [PubMed: 10207073]
71. Martowicz ML, Grass JA, Boyer ME, Guend H, Bresnick EH. Dynamic GATA factor interplay at a multi-component regulatory region of the GATA-2 locus. *J Biol Chem.* 2005; 280:1724–1732. [PubMed: 15494394]
72. Tripic T, Deng W, Cheng Y, et al. SCL and associated protein distinguish active from repressive GATA transcription factor complexes. *Blood.* 2008; 113:2191–2201. [PubMed: 19011221]
73. Mikkola HK, Klintman J, Yang H, et al. Hematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukemia SCL/tal-1 gene. *Nature.* 2003; 421:547–551. [PubMed: 12540851]
74. Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell.* 1996; 86:47–57. [PubMed: 8689686]
75. Hall MA, Slater NJ, Salmon JM, et al. Functional but abnormal adult erythropoiesis in the absence of the stem cell leukemia gene. *Mol Cell Biol.* 2005; 25:6355–6362. [PubMed: 16024775]
76. Palii CG, Perez-Iratxeta C, Yao Z, et al. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J.* 2011; 30:494–509. [PubMed: 21179004]
77. Gould KA, Bresnick EH. Sequence determinants of DNA binding by the hematopoietic helix-loop-helix transcription factor TAL1: importance of sequences flanking the E-box core. *Gene Expr.* 1998; 7:87–101. [PubMed: 9699481]

78. Wadman IA, Osada H, Grutz GG, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* 1997; 16:3145–3157. [PubMed: 9214632]
79. Kassouf MT, Chagraoui H, Vyas P, Porcher C. Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. *Blood.* 2008; 112:1056–1067. [PubMed: 18550854]
80. Dorsett D. Distant liaisons: long-range enhancer-promoter interactions in *Drosophila*. *Current Opinion in Genetics and Development.* 1999; 9:505–514. [PubMed: 10508687]
81. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science.* 2003; 302:415–419. [PubMed: 14564000]
82. Mylona A, Andrieu-Soler C, Thongjuea S, et al. Genome-wide analysis shows that Ldb1 controls essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis. *Blood.* 2013; 121:2902–2913. [PubMed: 23390196]
83. Li L, Jothi R, Cui K, et al. Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells. *Nat Immunol.* 2011; 12:129–136. [PubMed: 21186366]
84. Li L, Freudenberg J, Cui K, et al. Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood.* 2013; 121:4575–4585. [PubMed: 23610375]
85. Li L, Lee JY, Gross J, Song SH, Dean A, Love PE. A requirement for Lim domain binding protein 1 in erythropoiesis. *J Exp Med.* 2010; 207:2543–2550. [PubMed: 21041453]
86. Love PE, Warzecha C, Li L. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends in genetics: TIG.* 2014; 30:1–9. [PubMed: 24290192]
87. Miller IJ, Bieker JJ. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol.* 1993; 13:2776–2786. [PubMed: 7682653]
88. Pilon AM, Ajay SS, Kumar SA, et al. Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood.* 2011; 118:e139–148. [PubMed: 21900194]
89. Im H, Grass JA, Johnson KD, et al. Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. *Proc Natl Acad Sci U S A.* 2005; 102:17065–17070. [PubMed: 16286657]
90. Tallack MR, Magor GW, Dartigues B, et al. Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome Res.* 2012; 22:2385–2398. [PubMed: 22835905]
91. Tallack MR, Whittington T, Yuen WS, et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res.* 2010; 20:1052–1063. [PubMed: 20508144]
92. Tallack MR, Keys JR, Humbert PO, Perkins AC. EKLF/KLF1 controls cell cycle entry via direct regulation of E2f2. *J Biol Chem.* 2009; 284:20966–20974. [PubMed: 19457859]
93. Perkins AC, Sharpe AH, Orkin SH. Lethal beta-thalassaemia in mice lacking the erythroid CACCC- transcription factor EKLF. *Nature.* 1995; 375:318–322. [PubMed: 7753195]
94. Pilon AM, Arcasoy MO, Dressman HK, et al. Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol Cell Biol.* 2008; 28:7394–7401. [PubMed: 18852285]
95. Tsai FY, Keller G, Kuo FC, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature.* 1994; 371:221–226. [PubMed: 8078582]
96. Tsai F-Y, Orkin SH. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood.* 1997; 89:3636–3643. [PubMed: 9160668]
97. de Pater E, Kaimakis P, Vink CS, et al. Gata2 is required for HSC generation and survival. *J Exp Med.* 2013; 210:2843–2850. [PubMed: 24297996]
98. Dorfman DM, Wilson DB, Bruns GA, Orkin SH. Human transcription factor GATA-2. Evidence for regulation of preproendothelin-1 gene expression in endothelial cells. *J Biol Chem.* 1992; 267:1279–1285. [PubMed: 1370462]

99. Hattangadi SM, Wong P, Zhang L, Flygare J, Lodish HF. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*. 2011; 118:6258–6268. [PubMed: 21998215]
100. Weiss MJ, Keller G, Orkin SH. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev*. 1994; 8:1184–1197. [PubMed: 7926723]
101. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A*. 2003; 100:8811–8816. [PubMed: 12857954]
102. Wozniak RJ, Keles S, Lugus JJ, et al. Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. *Mol Cell Biol*. 2008; 28:6681–6694. [PubMed: 18779319]
103. Lugus JJ, Chung YS, Mills JC, et al. GATA2 functions at multiple steps in hemangioblast development and differentiation. *Development*. 2007; 134:393–405. [PubMed: 17166922]
104. Dore LC, Chlon TM, Brown CD, White KP, Crispino JD. Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. *Blood*. 2012; 119:3724–3733. [PubMed: 22383799]
105. May G, Soneji S, Tipping AJ, et al. Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. *Cell Stem Cell*. 2013; 13:754–768. [PubMed: 24120743]
106. Pal S, Cantor AB, Johnson KD, et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A*. 2004; 101:980–985. [PubMed: 14715908]
107. Sanalkumar R, Johnson KD, Gao X, et al. Mechanism governing a stem cell-generating cis-regulatory element. *Proc Natl Acad Sci U S A*. 2014
108. Groschel S, Sanders MA, Hoogenboezem R, et al. A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. *Cell*. 2014; 157:369–381. [PubMed: 24703711]
109. Yamazaki H, Suzuki M, Otsuki A, et al. A Remote GATA2 Hematopoietic Enhancer Drives Leukemogenesis in inv(3)(q21;q26) by Activating EVI1 Expression. *Cancer Cell*. 2014
110. Bender MA, Roach JN, Halow J, et al. Targeted deletion of 5'HS1 and 5'HS4 of the beta-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood*. 2001; 98:2022–2027. [PubMed: 11567985]
111. Bender MA, Bulger M, Close J, Groudine M. Beta-globin gene switching and DNaseI sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell*. 2000; 5:387–393. [PubMed: 10882079]
112. Hug BA, Wesselschmidt RL, Fiering S, et al. Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol*. 1996; 16:2906–2912. [PubMed: 8649401]
113. Fiering S, Epner E, Robinson K, et al. Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev*. 1995; 9:2203–2213. [PubMed: 7557375]
114. Bender MA, Mehaffey MG, Telling A, et al. Independent formation of DnaseI hypersensitive sites in the murine beta-globin locus control region. *Blood*. 2000; 95:3600–3604. [PubMed: 10828050]
115. Epner E, Reik A, Cimbara D, et al. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell*. 1998; 2:447–455. [PubMed: 9809066]
116. Sung YH, Baek IJ, Kim DH, et al. Knockout mice created by TALEN-mediated gene targeting. *Nature biotechnology*. 2013; 31:23–24.
117. Kim Y, Kweon J, Kim A, et al. A library of TAL effector nucleases spanning the human genome. *Nature biotechnology*. 2013; 31:251–258.
118. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*. 2013; 31:230–232.

119. Tijssen MR, Cvejic A, Joshi A, et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell*. 2011; 20:597–609. [PubMed: 21571218]
120. Wilson NK, Foster SD, Wang X, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 2010; 7:532–544. [PubMed: 20887958]
121. Beck D, Thoms JA, Perera D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 2013; 122:e12–22. [PubMed: 23974199]
122. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual review of biochemistry*. 2011; 80:437–471.
123. Linnemann AK, O'Geen H, Keles S, Farnham PJ, Bresnick EH. Genetic framework for GATA factor function in vascular biology. *Proc Natl Acad Sci U S A*. 2011; 108:13641–13646. [PubMed: 21808000]
124. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011; 118:2653–2655. [PubMed: 21670465]
125. Dickinson RE, Griffin H, Bigley V, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood*. 2011; 118:2656–2658. [PubMed: 21765025]
126. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*. 2011; 43:1012–1017. [PubMed: 21892162]
127. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet*. 2011; 43:929–931. [PubMed: 21892158]
128. Hsu AP, Johnson KD, Falcone EL, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood*. 2013; 121:3830–3837. S3831–3837. [PubMed: 23502222]
129. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science*. 1994; 265:1573–1577. [PubMed: 8079170]
130. Scott EW, Fisher RC, Olson MC, Kehrli EW, Simon MC, Singh H. PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. *Immunity*. 1997; 6:437–447. [PubMed: 9133423]
131. Leddin M, Perrod C, Hoogenkamp M, et al. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood*. 2011; 117:2827–2838. [PubMed: 21239694]
132. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38:576–589. [PubMed: 20513432]
133. Stopka T, Amanatullah DF, Papetti M, Skoultschi AI. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *EMBO J*. 2005; 24:3712–3723. [PubMed: 16222338]
134. Zhang PJ, Zhang X, Iwama A, et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood*. 2000; 96:2641–2648. [PubMed: 11023493]
135. Liu F, Kang I, Park C, et al. ER71 specifies Flk-1+ hemangiogenic mesoderm by inhibiting cardiac mesoderm and Wnt signaling. *Blood*. 2012; 119:3295–3305. [PubMed: 22343916]
136. Lee D, Park C, Lee H, et al. ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. *Cell Stem Cell*. 2008; 2:497–507. [PubMed: 18462699]
137. Sumanas S, Gomez G, Zhao Y, Park C, Choi K, Lin S. Interplay among Etsrp/ER71, Scl, and Alk8 signaling controls endothelial and myeloid cell formation. *Blood*. 2008; 111:4500–4510. [PubMed: 18270322]
138. Rasmussen TL, Kweon J, Diekmann MA, et al. ER71 directs mesodermal fate decisions during embryogenesis. *Development*. 2011; 138:4801–4812. [PubMed: 21989919]

139. Jochum W, Passegue E, Wagner EF. AP-1 in mouse development and tumorigenesis. *Oncogene*. 2001; 20:2401–2412. [PubMed: 11402336]
140. Kanki Y, Kohro T, Jiang S, et al. Epigenetically coordinated GATA2 binding is necessary for endothelium-specific endomucin expression. *EMBO J*. 2011; 30:2582–2595. [PubMed: 21666600]
141. Lichtinger M, Ingram R, Hannah R, et al. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J*. 2012; 31:4318–4333. [PubMed: 23064151]
142. Miyamoto T, Iwasaki H, Reizis B, et al. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev Cell*. 2002; 3:137–147. [PubMed: 12110174]
143. Akashi K. Lineage promiscuity and plasticity in hematopoietic development. *Ann N Y Acad Sci*. 2005; 1044:125–131. [PubMed: 15958705]
144. Hu M, Krause D, Greaves M, et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev*. 1997; 11:774–785. [PubMed: 9087431]
145. Cirillo LA, Zaret KS. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell*. 1999; 4:961–969. [PubMed: 10635321]
146. Bresnick EH, Martowicz ML, Pal S, Johnson KD. Developmental control via GATA factor interplay at chromatin domains. *J Cell Physiol*. 2005; 205:1–9. [PubMed: 15887235]
147. Kadauke S, Udugama MI, Pawlicki JM, et al. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell*. 2012; 150:725–737. [PubMed: 22901805]
148. Mouse EC, Stamatoyannopoulos JA, Snyder M, et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*. 2012; 13:418. [PubMed: 22889292]
149. Gerstein MB, Kundaje A, Hariharan M, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature*. 2012; 489:91–100. [PubMed: 22955619]
150. Ruau D, Ng FS, Wilson NK, et al. Building an ENCODE-style data compendium on a shoestring. *Nature methods*. 2013; 10:926. [PubMed: 24076986]
151. Hannah R, Joshi A, Wilson NK, Kinston S, Gottgens B. A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. *Exp Hematol*. 2011; 39:531–541. [PubMed: 21338655]
152. Chacon D, Beck D, Perera D, Wong JW, Pimanda JE. BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Res*. 2014; 42:D172–177. [PubMed: 24185696]
153. Shankaranarayanan P, Mendoza-Parra MA, Walia M, et al. Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nature methods*. 2011; 8:565–567. [PubMed: 21642965]
154. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature protocols*. 2011; 6:1656–1668.
155. Gilfillan GD, Hughes T, Sheng Y, et al. Limitations and possibilities of low cell number ChIP-seq. *BMC genomics*. 2012; 13:645. [PubMed: 23171294]
156. Moignard V, Macaulay IC, Swiers G, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol*. 2013; 15:363–372. [PubMed: 23524953]
157. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011; 144:296–309. [PubMed: 21241896]
158. Forsberg EC, Bresnick EH. Histone acetylation beyond promoters: long-range acetylation patterns in the chromatin world. *Bioessays*. 2001; 23:820–830. [PubMed: 11536294]
159. Wu CT, Morris JR. Transvection and other homology effects. *Curr Opin Genet Dev*. 1999; 9:237–246. [PubMed: 10322135]
160. Krivega I, Dean A. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev*. 2012; 22:79–85. [PubMed: 22169023]
161. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet*. 2013; 14:390–403. [PubMed: 23657480]

162. Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet.* 2005; 37:31–40. [PubMed: 15608638]
163. Handoko L, Xu H, Li G, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet.* 2011; 43:630–638. [PubMed: 21685913]
164. Fullwood MJ, Wei CL, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* 2009; 19:521–532. [PubMed: 19339662]

Figure 1.

Distinct functional properties of *Gata2* GATA switch sites. (A) ChIP-seq profiles of GATA-2 chromatin occupancy at *GATA2* locus in model erythroid cell systems: human K562 and mouse MEL cells. GATA-2 occupies established GATA factor binding regions termed GATA switch sites, which contain evolutionarily conserved GATA motifs and multiple attributes of transcriptional enhancers. (B) Schematic showing the GATA switch sites (–3.9, –2.8, –1.8 and +9.5 kb upstream of the 1S promoter transcription start site), and phenotypes resulting from deletion of the respective sites. The red X mark denotes the *cis*-element deletion from the mouse genome.