Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis

Ryan J. Wozniak  
*University of Wisconsin - Madison*

Sunduz Keles  
*University of Wisconsin - Madison*

Jesse J. Lugus  
*Washington University School of Medicine in St. Louis*

Ken H. Young  
*University of Wisconsin - Madison*

Meghan E. Boyer  
*University of Wisconsin - Madison*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

Please let us know how this document benefits you.

**Recommended Citation**

Wozniak, Ryan J.; Keles, Sunduz; Lugus, Jesse J.; Young, Ken H.; Boyer, Meghan E.; Tran, Tuan M.; Choi, Kyunghee; and Bresnick, Emery H., "Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis." Molecular and Cellular Biology. 28, 21. 6681–6694. (2008).  
[https://digitalcommons.wustl.edu/open_access_pubs/3112](https://digitalcommons.wustl.edu/open_access_pubs/3112)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Authors
Ryan J. Wozniak, Sunduz Keles, Jesse J. Lugus, Ken H. Young, Meghan E. Boyer, Tuan M. Tran, Kyunghee Choi, and Emery H. Bresnick
Molecular Hallmarks of Endogenous Chromatin Complexes Containing Master Regulators of Hematopoiesis

Ryan J. Wozniak, Sunduz Keles, Jesse J. Lugus, Ken H. Young, Meghan E. Boyer, Tuan M. Tran, Kyunghee Choi and Emery H. Bresnick

Published Ahead of Print 8 September 2008.

Updated information and services can be found at:
http://mcb.asm.org/content/28/21/6681

**SUPPLEMENTAL MATERIAL**
Supplemental material

**REFERENCES**
This article cites 94 articles, 58 of which can be accessed free at: http://mcb.asm.org/content/28/21/6681#ref-list-1

**CONTENT ALERTS**
Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
Molecular Hallmarks of Endogenous Chromatin Complexes Containing Master Regulators of Hematopoiesis

Ryan J. Wozniak,‡ Sunduz Keles,‡‡ Jesse J. Lugus, Ken H. Young, Meghan E. Boyer, Tuan M. Tran, Kyungee Choi, and Emery H. Bresnick

Departments of Pharmacology, Biostatistics and Medical Bioinformatics, Statistics, and Pathology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin 53706, and Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110

Received 8 July 2008/Returned for modification 5 August 2008/Accepted 28 August 2008

Combinatorial interactions among trans-acting factors establish transcriptional circuits that orchestrate cellular differentiation, survival, and development. Unlike circuits instigated by individual factors, efforts to identify gene ensembles controlled by multiple factors simultaneously are in their infancy. A paradigm has emerged in which the important regulators of hematopoiesis GATA-1 and GATA-2 function combinatorially with Scl/TAL1, another key regulator of hematopoiesis. The underlying mechanism appears to involve preferential assembly of a multimeric complex on a composite DNA element containing WGATAR and E-box motifs. Based on this paradigm, one would predict that GATA-2 and Scl/TAL1 would commonly co-occupy such composite elements in cells. However, chromosome-wide analyses indicated that the vast majority of conserved composite elements were occupied by neither GATA-2 nor Scl/TAL1. Intriguingly, the highly restricted set of GATA-2-occupied composite elements had characteristic molecular hallmarks, specifically Scl/TAL1 occupancy, a specific epigenetic signature, specific neighboring cis elements, and preferential enhancer activity in GATA-2-expressing cells. Genes near the GATA-2-Scl/TAL1-occupied composite elements were regulated by GATA-2 or GATA-1, and therefore these fundamental studies on combinatorial transcriptional mechanisms were also leveraged to discover novel GATA factor-mediated cell regulatory pathways.

Combinatorial interactions among trans-acting factors establish transcriptional circuits that control fundamental biological processes. In the context of metazoans, these interactions often occur at regulatory elements far from genes and within introns. Many genes require a complex collection of trans-acting factors, coregulator complexes, and long-range regulation, and therefore considerable challenges exist in forging general principles to explain combinatorial transcriptional control. We investigated combinatorial transcriptional mechanisms in the context of GATA factors, which interact with an assortment of regulatory factors to control differentiation, survival, and development (12, 42).

GATA-1 and GATA-2 have unique and essential roles to control hematopoiesis. GATA-2 is required for maintenance and expansion of hematopoietic stem cells (HSCs) (78, 79), while GATA-1 promotes the development of erythrocytes (20, 62, 63, 72), megakaryocytes (70), eosinophils (92), and mast cells (54). GATA-2 is also expressed in endothelial cells (17, 48, 56), and conditional GATA-2 expression in embryonic stem (ES) cells increases the genesis of hemangioblasts, precursors to hematopoietic and endothelial cells (50). GATA-2 deregulation is associated with early-onset coronary artery disease (15), atherosclerosis (69), and chronic myelogenous leukemia (94), whereas GATA-1 mutations cause megakaryoblastic leukemia (85) and additional blood disorders (16, 58).

Both GATA-1 and GATA-2 bind an identical DNA motif (WGATAR) (45, 52), but the majority of these motifs are unoccupied in cells (8, 26, 27, 34, 36, 51). Despite this shared binding specificity, GATA-1 and GATA-2 can exert distinct biological activities (21), indicating that each factor has certain unique targets and/or they differentially regulate common genes. GATA-1 and GATA-2 can occupy identical chromatin sites and induce opposite transcriptional outputs (9). However, they function redundantly to promote primitive erythroblast development (21). The modes by which GATA factors select target sites and mechanisms underlying their context-dependent functions are unresolved.

Context-dependent GATA-1 activity involves the capacity of GATA-1 to utilize diverse coregulators (5, 32, 68, 80) and the differential sensitivity of target loci to GATA factor levels (38). Combinatorial actions of GATA factors with other trans-acting factors are also important (42). A paradigm has emerged in which GATA-1 functions cooperatively with the E-box binding proteins Scl/TAL1 and E2A as well as LMO2 and LDB1 on WGATAR- and E-box (CANNTG)-containing composite elements in erythroid cells (47, 74, 83, 84, 89). In the context of naked DNA, these factors form a multimeric complex that preferentially recognizes such composite elements. Scl/TAL1 is expressed in GATA-1- and GATA-2-expressing hematopoietic cells (23, 24, 28), is induced by GATA-2 (13, 50), and is required for development of all hematopoietic cell types (66, 71), hematopoietic commitment of hemangioblasts (50), vasculogenesis (77), and angiogenesis (82).
Relative to GATA-1, considerably less is known about mechanisms underlying GATA-2 function. Only a few direct GATA-2 target genes are known, including genes encoding Scl/TAL1 (13, 50), GATA-2 itself (27, 46), and BMP4 (bone morphogenetic protein 4) (50). Although transcriptional elements uniquely controlled by GATA-2, but not other GATA factors, are unknown, an E-box–WGATAR composite element residing within a Gata2 intron (9.5 kb downstream of the transcription start site; hereafter referred to as the +9.5 kb site) confers strong enhancer activity in GATA-2-expressing cells in vitro and in the vasculature and fetal liver of mouse embryos (41, 88). The enhancer activity requires both WGATAR and E-box motifs (41, 88). Taken together with the paradigm that emerged from the finding that GATA-1 and Scl/TAL1 preferentially assemble a multimeric complex on composite elements in the context of naked DNA (84), one might predict that GATA-2 and Scl/TAL1 commonly co-occupy and function through such composite elements in vivo. However, chromosome-wide analyses revealed that the vast majority of conserved composite elements are not occupied by these factors. Mechanistic studies revealed specific molecular hallmarks that distinguished these unoccupied elements from a highly restricted subset of occupied composite elements. Furthermore, the occupied sites pinpointed novel GATA factor target genes that highlight new GATA factor-dependent cell regulatory pathways.

MATERIALS AND METHODS

Cell culture. GATA-1-null G1E cells resemble normal proerythroblasts, express endogenous GATA-2, and represent a powerful system for dissecting GATA factor mechanisms (26, 27, 29, 86, 87). G1E-ER–GATA-1 cells express an estrogen receptor ligand binding domain fusion to GATA-1, which interacts with endogenous GATA-2, and represent a powerful system for dissecting enhancer function in mouse embryos, the enhancer activity requires both WGATAR and E-box–WGATAR composite elements uniquely controlled by GATA-2, but not other GATA factors, are unknown, an E-box–WGATAR composite element residing within a Gata2 intron (9.5 kb downstream of the transcription start site; hereafter referred to as the +9.5 kb site) confers strong enhancer activity in GATA-2-expressing cells in vitro and in the vasculature and fetal liver of mouse embryos (41, 88). The enhancer activity requires both WGATAR and E-box motifs (41, 88). Taken together with the paradigm that emerged from the finding that GATA-1 and Scl/TAL1 preferentially assemble a multimeric complex on composite elements in the context of naked DNA (84), one might predict that GATA-2 and Scl/TAL1 commonly co-occupy and function through such composite elements in vivo. However, chromosome-wide analyses revealed that the vast majority of conserved composite elements are not occupied by these factors. Mechanistic studies revealed specific molecular hallmarks that distinguished these unoccupied elements from a highly restricted subset of occupied composite elements. Furthermore, the occupied sites pinpointed novel GATA factor target genes that highlight new GATA factor-dependent cell regulatory pathways.

RESULTS

Stringent molecular constraints for GATA-2-mediated combinatorial transcriptional control. The far upstream GATA-1- and GATA-2-binding region of the Gata2 locus (−77 kb) (27) resembles the +9.5 kb site in containing a conserved WGATAR and neighboring E-boxes (Fig. 1A). We tested whether the −77 and +9.5 kb sites function similarly in vivo using a LacZ vector identical to that used in our analysis of the +9.5 kb element (88).

The +9.5 kb vector contained a minimal Gata2 promoter fused to LacZ and was active in vascular endothelium, endocardium, and the fetal liver (88). Despite the common E-box and WGATAR motifs in both the +9.5 kb and the −77 kb elements, the −77 kb site failed to activate the Gata2 promoter-LacZ transgene in 12 out of 12 E11.5 F0 transgenic embryos (Fig. 1B). In addition to its enhancer function in mouse embryos, the +9.5 kb site activates a
Gata2 promoter-luciferase reporter in GATA-2-expressing endothelial (HUVEC) and hematopoietic (G1E) cells, which requires WGATAR and E-box motifs (88). Since the /H11002 77 kb site can activate the Gata2 promoter reporter in G1E cells, we tested whether it also functions in HUVECs. In contrast to the /H11001 9.5 kb site, the /H11002 77 kb site lacked activity (Fig. 1C).

The /H11001 9.5 kb site enhancer activity in endothelial cells requires an E-box–WGATAR-containing core module and regulatory modules containing additional cis elements (88), whereas the core module suffices for activity in GATA-2-expressing hematopoietic cells. The functional difference between +9.5 and −77 kb sites in HUVECs might therefore arise from the lack of −77 kb regulatory modules or differences in their core modules. To distinguish between these possibilities, chimeric elements were generated in which the WGATAR- and E-box-containing −77 kb core was substituted for the +9.5 kb core. The −77 kb core was incapable of reconstituting activity of the core-deleted +9.5 kb site, and a larger −77 kb core fragment containing the WGATAR motif and two flanking E-boxes also did not reconstitute activity (Fig. 1C). Thus, the −77 kb core differs from the +9.5 kb core, which confers GATA-2-dependent activation in endothelial cells.

Since the +9.5 and −77 kb cores are functionally distinct, presumably their cis-element compositions or configurations differ. The +9.5 kb core critically requires WGATAR and E-box motifs for activity in HUVECs and G1E cells (88). Although the −77 kb site lacks enhancer activity in HUVECs (Fig. 1C), it is active in G1E cells, and we tested whether this activity is WGATAR and E-box dependent. While mutation of WGATAR abrogated activity, mutation of the two E-boxes individually or collectively only slightly reduced activity (Fig. 1D). The −77 kb site activity therefore requires WGATAR, but the conserved E-boxes are largely unimportant in G1E cells. These results illustrate how WGATAR motifs suffice to mediate GATA factor function in certain contexts, while requiring additional cis-elements in other contexts.
FIG. 2. Strict architectural constraints for GATA factor-mediated combinatorial transcriptional control. (A) cis-element spacing requirements. Mutant plasmids were generated in which nucleotides between the E-box and WGATAR motifs were either deleted (−1, −2, and −3) or added.
activities and E-box utilization. Their WGATAR motifs have distinct orientations and spacing relative to neighboring E-boxes. To test whether spacing constraints exist for +9.5 kb activity in G1E cells and HUVECs, 1, 2, or 3 bp were deleted, and 1, 2, 3, 5, or 10 bp were inserted between the WGATAR motif and E-box. Whereas 1-bp deletions or insertions were tolerated with only 22 to 33% decreases in enhancer activity (Fig. 2A), deletions or insertions of ≥2 bp severely reduced activity. Scrambling 4 bp of intervening sequence (A4 core) did not affect activity, indicating that the specific intervening sequences are not essential. If the deletions or insertions inhibit activity by altering the helical orientation of the WGATAR and E-box with respect to each other, a 10-bp insertion that maintains the configuration should be inconsequential. However, the 10-bp insertion mutant lacked activity, consistent with a critical spacing constraint rather than a precise helical geometry.

To determine if the linear orientation of WGATAR relative to the E-box is important, the WGATAR and E-box orientations were reversed (Fig. 2B). While reversing the orientation of the E-box reduced enhancer activity by 38 and 46% in HUVEC and G1E cells, respectively, reversing the orientation of WGATAR, with or without 3 bp of flanking sequence, abrogated activity. Reversing the orientation of both WGATAR and the E-box was also devastating, whereas reversing the orientation of the full core and 3 bp of flanking sequence, which maintains the WGATAR orientation relative to the E-box, only slightly reduced activity. These results provide evidence for a stringent orientation constraint in which CANNTG resides upstream of WGATAR on the same strand of DNA. It is instructive to compare these findings to a site selection analysis with erythroleukemia cell extracts and randomized 26-bp oligonucleotides (84). This analysis used LMO2, Scl/TAL1, and E2A antibodies to select bound oligonucleotides containing an E-box 8 to 10 bp upstream of a GATA motif. Although the importance of spacing, helical geometry, and motif orientation on naked DNA binding, enhancer activity, and chromatin occupancy were not evaluated, our results on constraints for enhancer activity mirror those obtained from the site selection analysis. Thus, combinatorial regulation of enhancer activity with nonchromosomal templates might reflect GATA-2–E-protein nucleoprotein complex assembly, analogous to the proposed mechanism for GATA-1 (84).

Chromosome-wide analysis of GATA-2 occupancy at conserved composite elements; the vast majority of composite elements are unoccupied. Endogenous GATA-1 and an estrogen receptor ligand binding domain fusion to GATA-1 (ER–GATA-1) occupy a small percentage of WGATAR motifs in chromatin (8). Our studies at multiple loci revealed GATA-1 occupancy at a small subset (<10%) of conserved WGATAR motifs (8). GATA-1 and GATA-2 share many chromatin sites, but differences can exist (51). Whereas FOG-1 increases GATA factor occupancy is unknown. As E-box–WGATAR composite element function in GATA-2-expressing cells requires a precise geometry (Fig. 1 and 2), this geometry might facilitate GATA-2 chromatin occupancy or enhance GATA-2 function postoccupancy. We conducted quantitative ChIP analysis in GATA-1-null G1E cells to test whether GATA-2 preferentially occupies composite elements in a configuration that is optimal for enhancer activity (Fig. 2) versus WGATAR motifs lacking E-boxes within 20 bp of WGATAR. GATA-2 occupancy was analyzed at 63 conserved WGATAR motifs lacking E-boxes on chromosome 6 (Fig. 3A) and at all conserved composite elements on chromosomes 1, 6, and 7 (Fig. 3B to D). Amplicons encompassed WGATAR motifs or composite elements in which the WGATAR motif, E-box, and the intervening spacing, are conserved (mice to humans). To minimize gross differences in chromosomal positions and to ensure that ChIP signals did not overlap, we analyzed conserved WGATAR motifs lacking nearby E-boxes that were 3 to 50 kb from the corresponding composite elements on chromosome 6. GATA-2 occupied 4.8% (3/63) (Fig. 3A) and 9.8% (16/164) (Fig. 3B to E) of WGATAR motifs lacking E-boxes and composite elements, respectively. Statistical analysis using a z-test for two proportions indicated that the E-box does not significantly (P = 0.342) increase the probability of GATA-2 occupancy. The vast majority of both composite and WGATAR sites tested are unoccupied.

Diagnostic molecular hallmarks of GATA-2-occupied composite elements. Scl/TAL1 and GATA-1 preferentially assemble a complex on oligonucleotides containing E-box–WGATAR composite elements (84). As Scl/TAL1 is also expressed in GATA-2-expressing multipotent hematopoietic precursors (23, 28), we reasoned that Scl/TAL1 might reside at certain GATA-2-occupied composite elements. Quantitative ChIP analysis in G1E cells revealed little to no signal at the necdin and Eγ promoters that lack composite elements, at multiple conserved WGATAR motifs within the Gata2 and Fmod-Btg2 loci lacking E-boxes, and at conserved composite elements not occupied by GATA-2 (Fig. 4). In contrast, Scl/TAL1 occupied all GATA-2-occupied composite elements, and therefore Scl/TAL1 occupancy at the composite elements was absolutely predictive of GATA-2 occupancy.

Another potentially important hallmark of GATA-2-occupied composite elements is the local chromatin environment. The chromatin environment surrounding large numbers of WGATAR motifs or composite elements has not been described. Intriguingly, the occupied and unoccupied conserved composite elements had distinct epigenetic signatures. Epigenetic marks that often signify active chromatin, specifically di-acetylated H3, tetra-acetylated H4, and H3-dimeK4 (8),
FIG. 3. Chromosome-wide GATA-2 occupancy at conserved WGATAR motifs and E-box-WGATAR composite motifs. (A) Quantitative ChIP analysis of GATA-2 occupancy at 63 conserved WGATAR motifs (within 3 to 50 kb of the corresponding conserved composite motifs of panel B) across mouse chromosome (Chr.) 6 in G1E cells (mean ± standard error from three independent experiments). The numbers on the x axis refer to specific positions along the chromosome.
occupied composite elements (right). The GATA-2-occupied motifs are also shown at their specific chromosomal location. (E) Sequence composition of 148 unoccupied (left) and 16 experiments. The numbers on the x axis correspond to nearest-neighbor genes listed in Table S1 in the supplemental material. The numbers of GATA-2-occupied motifs are also shown at their specific chromosomal locations. (B to D) Quantitative ChIP analysis of GATA-2 occupancy at all conserved E-box–WGATAR composite elements and control sites lacking composite elements (mean ± standard error from three independent experiments).

FIG. 4. Scl/TAL1 occupancy at conserved composite elements occurs exclusively at GATA-2-occupied elements. Quantitative ChIP analysis was conducted in G1E cells to measure Scl/TAL1 occupancy at GATA-2-occupied, conserved composite elements, GATA-2-unoccupied, conserved composite elements, and control sites lacking composite elements (mean ± standard error from three independent experiments).

were selectively enriched at the occupied sites (Fig. 5). H3-dimeK36, whose distribution in functionally distinct chromatin regions is less well defined, was also enriched at the occupied sites (Fig. 5). In contrast, H3-trimeK9, which is often, but not always, present at repressed chromatin sites (8, 81), was selectively enriched at half of the unoccupied sites and was enriched at only 1 of 16 occupied sites. Considering the combinations of epigenetic marks, enrichments of acetylated H3 and H4, H3-dimeK4, and H3-dimeK36 combined with a deficiency of H3-trimeK9 are highly predictive of GATA-2 and Scl/TAL1 occupancy at conserved composite elements. As a sole predictive parameter, H3-trimeK9 was least useful, which might relate to the fact that this marker is enriched at certain repressed and active chromatin sites (81).

To reveal additional potential molecular hallmarks of GATA-2–Scl/TAL1 occupancy, rigorous statistical analysis was conducted to identify sequences that may discriminate between occupied and unoccupied composite elements. We tested whether sequences within the composite element correlate with occupancy by using a logistic regression model with occupancy status as the outcome and nucleotide compositions in each of the 12 degenerate positions (positions 3, 4, 7 to 15, and 20) within the 20 positions of the composite element as explanatory variables. This model also allowed two-way interactions of these positions. Neither variations among nucleotide compositions of the individual positions (Fig. 3E) nor their two-way interactions significantly correlated with occupancy.

Since multiple trans-acting factors interact with GATA factors (42) and certain factor (e.g., Ets) are required for GATA-2-dependent transcription in specific contexts (65), we tested whether their cognate motifs reside near the composite elements and correlate with occupancy. Consensus motifs for Ets factors, Sp1, EKLF, ZBP89, NF-E2/AP1, AML1/Runx1, and Gfi1b were identified in sequences flanking occupied and unoccupied composite elements (± 50, 100, 150, 250, or 500 bp from the composite element). We tested whether these sites discriminate between the two groups of composite elements. Only the Sp1 consensus [GT][GA]GGC[GT][GA][GA][GT] and 37/148 of the regions in occupied and unoccupied groups, respectively, have at least a copy of this motif within their flanks (P = 0.0016). Thus, the Sp1 consensus, which binds multiple factors (64), some of which interact with GATA-1 (53), and cooperates with WGATAR and additional cis-elements to establish Dnase I hypersensitivity (22), is significantly enriched in occupied versus unoccupied regions.

De novo analysis of sequences flanking the composite elements (± 50, 100, 150, 250, or 500 bp) using MEME (3) and COSMO (4) algorithms identified a highly significant differential enrichment of a novel motif [TC][CT][CT][TG][GT][GC] [CG][AT][GT][TG] in occupied versus unoccupied groups. This motif occurs within ±100 bp of the composite elements in 10/16 occupied regions and only 7/148 unoccupied regions (P = 4.109e−08), and has not been implicated in protein binding or transcription. The positions of the novel and Sp1 motifs with respect to occupied composite elements are shown

correspond to nearest-neighbor genes listed in Table S1 in the supplemental material. The numbers of GATA-2-occupied motifs are also shown at their specific chromosomal locations. (B to D) Quantitative ChIP analysis of GATA-2 occupancy at all conserved E-box–WGATAR composite motifs across mouse chromosome 6 (B), chromosome 1 (C), and chromosome 7 in G1E cells (mean ± standard error from three independent experiments). The numbers on the x axis correspond to nearest-neighbor genes listed in Table S1 in the supplemental material. The numbers of GATA-2-occupied motifs are also shown at their specific chromosomal location. (E) Sequence composition of 148 unoccupied (left) and 16 occupied composite elements (right). The x axis depicts the nucleotide position within the composite element, and the y axis represents the relative frequencies of the nucleotides.
in Fig. S1 in the supplemental material. A combinatorial rule, in which the Sp1 and/or the novel motif resides at a locus, identifies 14/16 occupied sites as occupied and only 21/148 unoccupied sites as occupied ($P = 5.596 \times 10^{-8}$). Neither the novel nor the Sp1 motif significantly associates with occupied WGATAR sequences lacking nearby E-boxes. Collectively, the Scl/TAL1 occupancy, the epigenetic patterns, and the statistical distribution of specific cis elements establish diagnostic molecular hallmarks of Scl/TAL1–GATA-2 complexes on composite elements in hematopoietic cells.
Based on the molecular hallmarks described above, we reasoned that Scl/TAL1–GATA-2 complexes assembled at composite elements would have the capacity to regulate transcription. To test whether occupied composite elements function as GATA factor-dependent enhancers, we analyzed the activities of the respective elements with 10 bp of flanking sequence on the 5' and 3' ends in a transient transfection assay (Fig. 6A). While 7 out of 10 occupied composite elements conferred statistically significant enhancer activity in GATA-2-expressing G1E cells, with three exhibiting particularly notable activity, only 1 out of 10 composite elements (Gata2 +9.5 element) had significant activity in GATA-1-expressing MEL cells, and this activity was low. Similar results were obtained in DMSO-induced MEL cells (data not shown). In contrast, the ~3.9 kb region of the Gata2 locus, which lacks a composite element but is activated by GATA-1 via a WGATAR motif (51), was highly active in MEL cells. Thus, a subset of the occupied composite elements function as enhancers, preferentially in GATA-2-expressing cells.
which these motifs were mutated (Fig. 6B). Mutation of the E-box, WGATAR, or both motifs abrogated enhancer activity (Fig. 6B).

**Novel cell regulatory pathways derived from analysis of chromatin targets.** As GATA-2 occupied composite elements near genes not known to be GATA factor target genes, we tested whether Dox-mediated induction of GATA-2 during mouse ES cell differentiation into EBs (50) alters expression of genes neighboring the occupied composite elements (Fig. 7A). We also compared gene expression in Gata2<sup>+/−</sup> and wild-type EBs. Addition of Dox on day 2 of the culture greatly increased Gata2 mRNA in day 3, 4, 6, and 8 EBs derived from iGATA-2 ES cells (Fig. 7B). Similarly, Dox significantly induced the established GATA-2 target (50) Scl/TAL1 (12-fold), and the novel GATA-2 targets Sox6 and Fmod (4.1- and 68-fold, respectively) (Fig. 7B). Sox6 promotes chondrogenesis (73), represses embryonic β-like globin transcription in definitive erythroid cells (91), and functions cell autonomously to promote proliferation, survival, and differentiation of erythroid cells (18). Fibromodulin is a small, leucine-rich proteoglycan that suppresses signaling mediated by transforming growth factor β (TGF-β) and related cytokines (90), and the related protein biglycan binds BMP4, suppressing BMP4 signaling (57). Dox significantly decreased Btg2 expression (2.1-fold), while expression of other genes changed <2-fold. Btg2 (1) functions downstream of p53 to promote cell cycle arrest (6), suppresses Ras-induced transformation (6), and is deregulated in breast and renal cancer (40, 76).

_Fmod_ was highly downregulated in Gata2-null EBs (Fig. 7C). Taken together with the striking Dox-dependent _Fmod_ induction (68-fold) and the endogenous GATA-2 interaction with the _Fmod_ locus, these results establish _Fmod_ as a bona fide GATA-2 target gene (Fig. 7D). The magnitude of the GATA-2-mediated _Fmod_ induction is considerably greater than that for any other GATA-2 target gene reported. _Fmod_ is the most highly overexpressed gene in B-cell chronic lymphocytic leukemia (CLL) (35, 44, 55), but mechanisms underlying its normal transcriptional control and deregulation in CLL are unknown. Fibromodulin is secreted from CLL cells and resides intracellularly (55). As fibromodulin suppresses signaling by TGF-β and related proteins (90), biglycan suppresses BMP4 signaling (57), proteoglycans suppress signaling by BMP factors in _Drosophila melanogaster_ (7), and BMP4 induces _Gata2_ expression (19, 50), _Fmod_ overexpression in CLL might disrupt a circuit (Fig. 7D) in which BMP4 induces and/or sustains _Gata2_ expression in the hematopoietic niche. The role of these factors in hematopoiesis and leukemogenesis is unexplored.

Expression of the additional genes in wild-type versus _Gata2<sup>−/−</sup>_ EBs differed by less than twofold (Fig. 7C), although it cannot be ruled out that _GATA-2_ regulates these genes redundantly with other factors and/or nonredundantly in distinct contexts. We also conducted complementation analysis in _GATA-1_-null proerythroblast-like G1E cells stably expressing ER-ERE-GATA-1 (26, 27, 29, 34, 43, 51, 60), β-Estradiol-mediated ER-GATA-1 activation instigates GATA-2 displacement at chromatin sites, and “GATA switches” increase or decrease transcription (9). ER-GATA-1-mediated activation or repression can result from disruption of GATA-2-mediated transcriptional control. ER-GATA-1 activation in G1E-ERE- _GATA-1_ cells repressed _Gata2_ (7b-fold) (Fig. 7B, right), as expected (26), and also repressed _Sox6, Tram2_, and _Etv6_ (9.1-, 21-, and 3.2-fold, respectively). _Tram2_ is a BMP and Runx2 target gene in osteoblasts (67) and regulates protein translation in the endoplasmic reticulum (75). _Etv6_ (Tel), a member of the _Ets_ transcription factor family, is a key regulator of adult HSCs that is frequently disrupted via leukemogenic chromosomal translocations (30, 93). ER-GATA-1 induced _Btg2_, _Bpgm_, _Foxp1_, and _Bcl2l13_ (323-, 15-, 8.5-, and 6.0-fold, respectively). _Bpgm_ encodes bisphosphoglycerate mutase, which catalyzes synthesis of the major allosteric regulator of hemoglobin (11). The Forkhead transcription factor _Foxp1_ is a prognostic factor in diffuse large B-cell lymphoma (10), but whether it functions in hematopoiesis is unclear. The proapoptotic factor _Bcl2l13_ (39) is a prognostic factor in B-lineage acute lymphocytic leukemia (31). _Fmod_ was expressed at almost undetectable levels in untreated and estradiol-treated G1E-ERE-GATA-1 cells, and considering the ES cell data, _GATA-2_ is necessary but insufficient to confer high-level _Fmod_ expression.

**DISCUSSION**

We have demonstrated that GATA-2 occupies a highly restricted subset of conserved composite elements, and this subset exhibited unique molecular hallmarks relative to the highly abundant unoccupied elements. GATA-2-occupied sites were co-occupied by Scl/TAL1 and had both a distinctive epigenetic signature and specific neighboring cis elements. These parameters offer considerable predictive value and provide important mechanistic insights vis-a-vis GATA-2 recognition of target sites in chromatin.

The GATA-2-Scl/TAL1 occupancy at all GATA-2-occupied composite elements strongly suggests that these factors function combinatorially at common targets in the genome. As our quantitative comparison of GATA-2 occupancy at composite elements versus isolated WGATAR motifs did not reveal significant differences, it seems unlikely that the composite element serves to increase the probability of GATA-2 occupancy. Thus, it is attractive to propose that both factors function collectively post-chromatin occupancy to recruit requisite co-regulators, thereby conferring combinatorial transcriptional control.

Since GATA-1 functionally interacts with Scl/TAL1 in the context of composite elements (84), we assumed that the GATA-2- and Scl/TAL1-occupied composite elements would function as GATA-1- and perhaps GATA-2-responsive enhancers. However, while a subset of these elements functioned as enhancers in GATA-2-expressing cells, 9 out of 10 were not significantly active in GATA-1-expressing cells (Fig. 6A). Thus, it is instructive to compare our results with those of studies implicating GATA-1 in functioning through DNA sequences containing E-box and WGATAR motifs. In gel-shift assays using MEL cell nuclear extracts, GATA-1, the E-proteins Scl/TAL1 and E2A, and their interacting proteins Lmo2 and Ldb1 assemble a complex on oligonucleotides containing E-box and WGATAR elements. These parameters offer considerable predictive value and provide important mechanistic insights vis-a-vis GATA-2 recognition of target sites in chromatin.

Since GATA-1 functionally interacts with Scl/TAL1 in the context of composite elements (84), we assumed that the GATA-2- and Scl/TAL1-occupied composite elements would function as GATA-1- and perhaps GATA-2-responsive enhancers. However, while a subset of these elements functioned as enhancers in GATA-2-expressing cells, 9 out of 10 were not significantly active in GATA-1-expressing cells (Fig. 6A). Thus, it is instructive to compare our results with those of studies implicating GATA-1 in functioning through DNA sequences containing E-box and WGATAR motifs. In gel-shift assays using MEL cell nuclear extracts, GATA-1, the E-proteins Scl/TAL1 and E2A, and their interacting proteins Lmo2 and Ldb1 assemble a complex on oligonucleotides containing E-box and WGATAR elements. These parameters offer considerable predictive value and provide important mechanistic insights vis-a-vis GATA-2 recognition of target sites in chromatin.
FIG. 7. Occupied composite elements reside at and near novel GATA factor target genes. (A) Diagrams of nearest-neighbor genes surrounding GATA-2-occupied E-box–GATA motifs on mouse chromosomes (Chr.) 1, 6, and 7. Asterisks denote the locations of conserved E-box–WGATAR motifs, arrows denote transcription start sites, and shaded boxes indicate the coding regions of the genes. (B) The table summarizes changes (fold) in the GAPDH-normalized expression of selected genes surrounding GATA-2-occupied E-box–WGATAR composite elements in day 3/4, 6, and 8 EBs derived from mouse ES cells following Dox-mediated GATA-2 induction (mean ± standard error from nine independent experiments for day 3/4 and 6 EBs and from six independent experiments for day 8 EBs) and also in G1E-ER–GATA-1 cells after estradiol-mediated activation of ER–GATA-1 (mean ± standard error from three independent experiments). mRNA levels were quantitated by real-time RT-PCR. (C) The graph depicts the GAPDH-normalized expression of genes surrounding GATA factor-occupied E-box–WGATAR composite elements in day 3 and 6 EBs derived from Gata2−/− ES cells divided by their expression in day 3 and 6 EBs derived from wild-type ES cells, respectively (mean ± standard error from three independent experiments). mRNA levels were quantitated by real-time RT-PCR. (D) Model of BMP4–GATA-2–fibromodulin regulatory circuit.
elements, functions as an enhancer in MEL cells (89). Furthermore, smaller sequences containing an E-box–E-box–WGATAR motif (14) or a WGATAR–E-box–WGATAR motif (2) function as enhancers in MEL cells (approximately twofold and approximately fivefold, respectively). A 1.1-kb Gata1 regulatory element (HS1) containing a WGATAR motif 9 bp downstream from an E-box activates LacZ expression in primitive erythroid cells in the yolk sac at E10.5 and in definitive erythroid cells in the fetal liver of E14.5 transgenic mouse embryos (83). In contrast, a smaller HS1 derivative (62 bp) containing only the WGATAR motif without the E-box retains activity, and therefore the E-box is not required for activity of this sequence in erythroid cells. Our study, which represents the first to analyze E-box–GATA composite motifs with defined spacing/orientation, demonstrates striking differences in their responsiveness to GATA-1 versus GATA-2 and also different intrinsic activities in GATA-2-expressing cells. Thus, it is attractive to propose that cis elements neighboring the composite elements influence their efficacy as enhancers and also specificity vis-à-vis different GATA factors.

Intriguingly, a specific epigenetic signature was one of the molecular hallmarks deduced that distinguishes occupied versus unoccupied composite elements. This specific chromatin modification state might be a prerequisite for GATA-2–Sel/TAL1 complex assembly at composite elements in hematopoiesis, or the complex might establish this signature post-chromatin occupancy. Since GATA-2 is expressed in HSCs, and no trackable systems exist to examine chromatin structure at target sites prior to the physiological appearance of GATA-2, distinguishing between these models will be challenging. In principle, one could ask whether knocking down GATA-2 in G1E cells reconfigures the epigenetic signature. However, we have not achieved efficacious GATA-2 knockdowns in this system. Even if a satisfactory degree of knockdown can be achieved, persistence of the epigenetic signature might reflect its establishment prior to complex assembly or a GATA-2 requirement for establishment, but not maintenance, of the signature. Nevertheless, our results provide the first example in which GATA-2-occupied target sites, or target sites occupied by any GATA factor or master regulator of hematopoiesis, exhibit a specific epigenetic signature that provides a unique foundation for conducting subsequent mechanistic analysis.

The GATA-2–Sel/TAL1 co-occupied target sites pinpointed neighboring genes that are GATA-2 and/or GATA-1 responsive. Thus, the functional genomic strategy established novel links between GATA-2 and proteins involved in important biological processes, including leukemogenesis. Other than Btg2, which was described in an analysis of GATA-1 target genes (87), GATA factors had not been linked to the regulation of genes identified in our screen. Furthermore, no prior reports have identified GATA-2 target genes using the multiple approaches described herein: endogenous GATA-2 occupancy of endogenous loci, regulation of the endogenous genes upon conditional expression of GATA-2 in physiologically relevant ES cells, and also regulation of the endogenous genes in Gata2 knockout cells. The novel targets (Fig. 7B) define five transcriptional modes: (i) GATA-2 activated, (ii) GATA-2 activated and GATA-1 repressed, (iii) GATA-1 activated, (iv) GATA-1 repressed, and (v) neither GATA-1 nor GATA-2 regulated. It will be particularly instructive to assemble the genome-wide genetic network collectively instigated by GATA-2 and Sel/TAL1 at composite elements to gain a comprehensive view of how stem cells give rise to diverse lineages of blood cells, how perturbations of this network causes leukemias, and, more fundamentally, how two cell-type-specific trans-acting factors selectively recognize their targets and function at a highly restricted subset of composite elements in a complex genome.

ACKNOWLEDGMENTS

We thank Stuart Orkin for providing Gata1−/− ES cells. We appreciate critical comments from the Bresnick group.

This work was funded by NIH grants DK06834 (E.H.B.), HG003747 (S.K.), and HL55337 (K.C.); National Research Service Award T32 HL 07936 from the NHLBI to the University of Wisconsin—Madison Cardiovascular Research Center (R.J.W.); Ruth L. Kirschstein fellowship F32 HL092736 from the National Heart, Lung, and Blood Institute (R.J.W.); and the PhRMA Foundation Research Starter Grant in Informatics (S.K.).

The content of this article does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

REFERENCES


