Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression

Timothy M. Chlon,[1,2] Maureen McNulty,[1] Benjamin Goldenson,[1] Alexander Rosinski,[1] and John D. Crispino[1]

[1]Division of Hematology/Oncology, Northwestern University, Chicago, IL, USA; and [2]*Present address Cincinnati Children's Hospital, Cincinnati, OH, USA*

ABSTRACT

GATA1 is a master transcriptional regulator of the differentiation of several related myeloid blood cell types, including erythrocytes and megakaryocytes. Germ-line mutations that cause loss of full length GATA1, but allow for expression of the short isoform (GATA1s), are associated with defective erythropoiesis in a subset of patients with Diamond Blackfan Anemia. Despite extensive studies of GATA1s in megakaryopoiesis, the mechanism by which GATA1s fails to support normal erythropoiesis is not understood. In this study, we used global gene expression and chromatin occupancy analysis to compare the transcriptional activity of GATA1s to GATA1. We discovered that compared to GATA1, GATA1s is less able to activate the erythroid gene expression program and terminal differentiation in cells with dual erythroid-megakaryocytic differentiation potential. Moreover, we found that GATA1s bound to many of its erythroid-specific target genes less efficiently than full length GATA1. These results suggest that the impaired ability of GATA1s to promote erythropoiesis in DBA may be caused by failure to occupy erythroid-specific gene regulatory elements.

Introduction

Megakaryocytes and erythroid cells are derived from a common progenitor cell termed the megakaryocyte-erythrocyte progenitor (MEP), and the differentiation of both of these cell types depends on the zinc finger transcription factor GATA1. As a master transcriptional regulator, GATA1 directs global gene expression to specify terminal erythroid or megakaryocytic fate.[1-5] It accomplishes this by both activating lineage specifying genes and repressing progenitor maintenance genes through a variety of gene regulatory mechanisms that depend on context-dependent co-factor interactions.[6] Importantly, GATA1 interacts with and controls the expression of many other lineage-specifying transcription factors to co-ordinately repress factors promoting other cell fates while activating those of megakaryocytes and erythrocytes.[7,8]

The lineage fate decision of the MEP towards either the erythroid or megakaryocyte fate is controlled by complex interactions among transcription factors.[7] To specify megakaryocytes, GATA1 co-operates with several ETS family transcription factors, including FLI1, ETS2, and ERG, to bind and activate megakaryocyte-specific genes.[9-11] In erythroid cells, GATA1 activates the Kruppel family transcription factor KLF1, which binds and activates erythroid genes.[12-18] However, the mechanism by which GATA1 contributes to lineage specification is poorly understood.

GATA1 exists as two isoforms in human cells, full-length protein and a shorter isoform named GATA1s, which is expressed from a downstream alternative start site (Met84).[19] GATA1s lacks the first 83 amino acids which comprise the N-terminal transactivation domain. Of note, the molecular mechanism by which this domain contributes to gene regulation is not well understood.[20-22] Acquired *GATA1* mutations that cause loss of full-length protein, and thus expression of only GATA1s, are involved in the pathogenesis of both transient myeloproliferative disorder and acute megakaryocytic leukemia (AMKL) in children with Down syndrome (DS).[23-25] Remarkably, *GATA1* mutations are detectable in nearly 30% of infants with DS.[26] Additional mutations that contribute to AMKL have recently been discovered. These include mutations in genes that control signaling pathways, including *JAK2*, *MPL*, and *RAS*, epigenetic regulators, such as *EZH2*, *CTCF*, and the cohesin complex.[27,28]

Germ-line *GATA1* mutations that lead to the exclusive production of GATA1s in the absence of DS have also been described. In one family, an inherited GATA1s mutation was found associated with impaired erythropoiesis and irregularities in the megakaryocyte lineage.[29] More recent studies have discovered GATA1s mutations in a subset of Diamond Blackfan Anemia (DBA) patients who lack mutations in ribosomal genes.[30,31] Together these studies suggest that GATA1s cannot support normal erythropoiesis.

In order to characterize the transcriptional activity of GATA1s in erythroid and megakaryocytic development, we completed global gene expression and chromatin occupancy analysis of GATA1 and GATA1s in a cell line with dual erythroid and megakaryocytic differentiation potential. Our data reveal that GATA1s binds and activates megakaryocyte-specific genes normally. In contrast, it fails to bind and activate erythroid genes to the same extent as GATA1. This deficiency of GATA1s in DNA binding and gene expression is associated with an impaired ability to promote development of erythroid cells. Together, these findings suggest that the activation domain of GATA1 is required for gene expression and differentiation of erythroid progenitor cells.

 doi:10.3324/haematol.2014.112714
Manuscript received on June 25, 2014. Manuscript accepted on February 2, 2015.
Correspondence: j-crispino@northwestern.edu

Methods

Cell culture

G1ME cells were maintained in 1% THPO-conditioned media or in media containing THPO, SCF and EPO, as described.[3]

Retroviral transduction and constructs

Retroviral supernatant was prepared and applied to cells, as previously described.[32] MigR1 constructs containing HA-tagged GATA1 and GATA1s have been described previously.[22]

Cell sorting and flow cytometry

GFP+ cells were purified with a MoFlo cell sorter (Beckman Coulter). Staining for CD42, Ter119, and DNA content was performed as previously described.[22] Antibodies for FACS analysis included anti-CD42-DyLight649 (Emfret, M040-3) and anti-Ter119 (BD Pharmingen, 561071).

Antibodies

GATA1 (sc-1234) and HA-tag (sc-7392) were purchased from Santa Cruz Biotechnology.

ChIP-seq

ChIP was performed as previously described.[33] For ChIP-Seq, 50E6 cells were infected with MigR1 viruses and harvested after 48 h. Three biological replicate anti-HA tag ChIP samples and 3 input samples were processed as previously described[34] and sequenced on a Genome Analyzer II or IIx (Illumina). Sequence tags were mapped to the mouse genome (mm9) using the Illumina pipeline.

ChIP-Seq binding site identification

To identify ChIP-Seq peaks, the peak calling program QuEST was used on default parameters.[35] To identify GATA1s-deficient sites, the threshold score for enrichment on QuEST was reduced from 30 to 10, and the resulting peak set was overlapped with the original GATA1 set using the ChIP-Seq Tool Kit.[36] GATA1 peaks without a peak in the expanded GATA1s peak set were identified as GATA1s-deficient. For analysis of genomic location and overlap with gene expression data, each peak was assigned to the gene with the nearest TSS using the ChIP-Seq Tool Set.[36] Control peak sets were generated using a custom Perl script described previously.[37] ChIP-seq data are available at GEO (accession # GSE64327).

Motif analysis

MEME[38] was used for *de novo* motif finding. 200bp of genomic sequence surrounding each peak summit was submitted to MEME using the 'zoops' option and max motif length set at 10 bases. For informed motif finding, the Homer findmotifs program was used on default parameters in conjunction with the provided set of known motifs.[39]

Gene expression analysis

G1ME cells were transduced with MigR1 retroviruses and the GFP+ cells were sorted 68 h later. The cells were allowed to recover in culture for 4 h, and then the total RNA was extracted using RNeasy Plus mini columns (Qiagen) following the manufacturer's protocol. The samples were hybridized to Illumina MouseWG-6 v.2.0 Expression BeadChips. The background was subtracted and the data werew quantile normalized to remove batch effects. The probe level data were analyzed using GeneSpring software (Agilent Technologies) as described previously.[5] Heatmaps were generated using the software program Gene Cluster 3.0,[40] and then the Java program TreeView was used to make and edit the heatmap.[41] Microarray data are available at GEO (accession # GSE64496).

Statistical analysis

For quantitative assays, treatment groups were reported as mean±SD and compared using the unpaired Student t-test. To test for independence of groups based on categorical data, χ^2 analysis with the Yates Correction for large sample sizes was used. $P \leq 0.05$ was considered statistically significant.

Table 1. Genes that show reduced GATA1s binding and are differentially expressed by GATA1s relative to GATA1.

Gene Symbol	Fold-change	Direction
Alas2	41.771973	down
Slc4a1	24.076063	down
Pdia2	9.113187	down
Shank3	8.295997	down
Kctd7	6.723526	down
Zfp236	6.147942	down
Bace2	6.02214	down
Cldn13	5.711454	down
Lrrfip1	5.2232666	down
Bmp2k	4.8215065	down
Ampd3	4.0424047	down
Klf1	3.4269867	down
Zbtb46	3.3146002	down
Jub	3.1339316	down
Gucy2g	2.431713	down
Lmo2	2.10571	down
Kcnn4	2.037378	down
Alad	1.9424441	down
Tarsl2	1.9198862	down
1110032E23Rik	1.9192394	down
Pnma1	1.8863618	down
Pbx1	1.8386884	down
Tnrc6b	1.8221127	down
Iars	1.76903	down
Slc14a1	1.6897564	down
Ptdss2	1.6818486	down
Fbxl10	1.64049	down
Slc43a1	1.6322525	down
Tiam1	1.612428	down
2010011I20Rik	1.5481253	down
Mbp	1.5046816	down
Cox4i2	9.756168	up
Mylip	4.372835	up
Snx5	3.988198	up
Cecr5	3.3212414	up
Pde4dip	3.048113	up
Sox6	2.874852	up
Prdm10	2.6762817	up
Tgm2	2.2664144	up
B230342M21Rik	1.7526656	up
Tnfrsf13b	1.6281554	up
Mbd1	1.5634671	up

Results

GATA1 and GATA1s induce similar degree of megakaryocyte differentiation

G1ME is a cell line that was derived from *Gata1*–deficient ES cells and has both megakaryocytic and erythroid differentiation potential upon reconstitution of GATA1 expression.[3] This gene complementation system provides a controlled cell context in which to study the activities of mutants of *GATA1*. We infected G1ME cells cultured in thrombopoietin (THPO) with MigR1 retroviruses expressing HA-tagged GATA1 or GATA1s and monitored megakaryocytic differentiation over six days (Figure 1A). GATA1 induced megakaryocyte maturation, including increased cell size, polyploidy, and expression of the cell surface marker CD42 (Figure 1B). The percentage of GATA1-transduced cells, marked by GFP expression, decreased rapidly and the transduced cells disappeared by day 6 (Figure 1C). This finding suggests that the GATA1 expressing cells fully differentiated or died within six days. In comparison, we observed that GATA1s also efficiently induced megakaryocytic differentiation, but some subtle differences were seen. For example, at days 3 and 4, GATA1s cultures had a significantly higher percentage of GFP$^+$ cells, suggesting that there is either delayed apoptosis, or an uncoupling of proliferation arrest and differentiation due to the loss of the N-terminus (Figure 1C). Similar percentages of CD42$^+$ cells were observed in the cultures, indicating GATA1s is not defective in inducing commitment to the megakaryocytic lineage. However, decreased levels of polyploidy were observed in GATA1s cultures at days 3 and 4, suggesting that the GATA1s cells were modestly delayed in terminal differentiation (Figure 1C). By western blot analysis, with both anti-HA-tag and anti-GATA1 antibodies, we found that GATA1s was expressed at much higher levels than GATA1 (Figure 1D). Though we are uncertain as to why GATA1s was expressed at higher levels than GATA1, elevated level of GATA1 associated with megakaryocytic differentiation is consistent with the hypomorphic activity of GATA1s observed in animal models.[42]

GATA1s is less efficient than GATA1 at promoting erythroid cell differentiation

Next, we sought to determine to what extent the G1ME cells differentiate towards the erythroid lineage upon reconstitution with GATA1 or GATA1s. Thus, we examined the expression of the erythroid marker Ter119 by flow cytometry at 72 h post transduction with GATA1 and GATA1s (Figure 2A). We observed that, on average, 8% of GATA1-expressing cells exhibited Ter119-positivity, whereas only 4% of the GATA1s-expressing cells stained for Ter119. We repeated this study in G1ME cells that were cultured with EPO, SCF, and THPO. Under similar conditions, a previous study found that with expression of GATA1, 59% and 22% cells of the differentiated into megakaryocytes (CD42$^+$) and erythroid cells (Ter119$^+$), respectively.[3] In our hands, expression of full length GATA1 led to, on average, 62% megakaryocytes and 27% erythroid cells (Figure 2B and C). Similar to what we observed in cells cultured with THPO alone, GATA1s showed an impaired ability to produce erythroid cells, with an average decline of 50%, but an additional propensity to produce more megakaryocytes (Figure 2B and C).

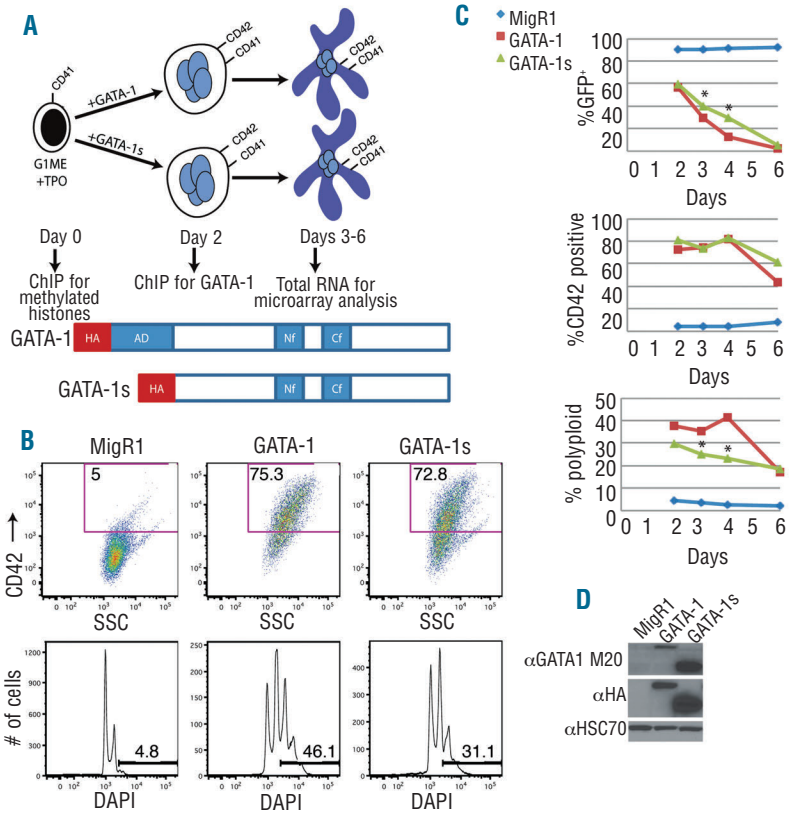

Figure 1. GATA1 and GATA1s induce megakaryocytic differentiation of G1ME cells. (A) A schematic presentation of the induction of megakaryocyte differentiation of G1ME by retroviral transduction with GATA1 or GATA1s. (B) FACS analysis of CD42 expression and DNA content of G1ME cells 72-h post transduction. (C) The percentages of GFP$^+$ cells, CD42$^+$ cells, and polyploidy (>4N) cells at 2-6 days post transduction with control, GATA1, or GATA1s virus. Graphs show mean±SD, n=3; *P<0.05. (D) Western blot of nuclear lysates from G1ME cells 48-h post transduction with empty vector, GATA1, or GATA1s.

We then examined the red cells derived from G1ME cells in more detail. Benzidine staining of cytospins confirmed that there was a reduction in production of erythroid cells in GATA1s *versus* GATA1 cultures (Figure 2D). Finally, qRT-PCR analysis of sorted Ter119^{+} cells revealed that although GATA1s reconstituted cells expressed *Hbb*, *Klf1*, and *Ank1*, the levels were significantly lower in GATA1s red cells than that observed in GATA1 reconstituted cells (*data not shown*). Taken together, we conclude that GATA1s induces megakaryocyte development similar to GATA1, but is less able to activate the erythroid program.

ChIP-seq analysis of GATA1 and GATA1s in G1ME cells reveals differences in chromatin binding

To compare the genome-wide transcriptional activity between GATA1s and full length GATA1 during megakaryopoiesis, we utilized ChIP-Seq and microarray gene expression profiling of G1ME cells. We harvested G1ME cells at 48 h post transduction with GATA1 or GATA1s, and then performed ChIP using an anti-HA antibody. We confirmed that the ChIP enriched both GATA1 and GATA1s bound chromatin by qPCR for selected known GATA1 binding sites. We found both proteins were enriched similarly at the promoter of *Pf4*, a megakaryocyte-specific GATA1-target gene, and the intron of *Hhex*, a homeobox gene that is expressed broadly in hematopoietic progenitor cells and is repressed by GATA1 (Figure 3A and B). We then performed three replicate ChIP experiments and sequenced the DNA by high-throughput next generation sequencing. Together, these experiments generated 20.1E6 and 20.5E6 total mappable reads for GATA1 and GATA1s, respectively. Using the QuEST peak calling program set on default parameters, these data yielded 2728 GATA1 peaks, but only 979 GATA1s peaks. The basepair-wise overlap of these datasets revealed that 853 peaks are shared while there are 1875 peaks specific to GATA1 and just 126 peaks specific to GATA1s (Figure 3C). Next, using *de novo* motif finding analysis, we found a canonical GATA (WGATAR) site in 84% of the GATA1 peaks and 73% of the GATA1s peaks, confirming the specificity of the analysis (Figure 3D). We also observed strong enrichment of binding motifs for ETS family transcription factors, which are known to co-operate with GATA1 in megakaryocytic differentiation (60% and 52%).[4,43] We then mapped each peak location to the closest transcription start site and assigned the peaks to the categories: upstream (2-100kb), promoters (0-2kb), intragenic, downstream (0-100kb), and gene desert (>100kb). This analysis revealed that the GATA1 peaks were significantly enriched at promoters and intragenic regions and significantly depleted in gene desert regions compared to randomly distributed genomic regions, consistent with the expected locations for gene regulatory regions (Figure 3E). Despite differences in the number of occupied regions, the GATA1s peaks had a similar distribution as GATA1.

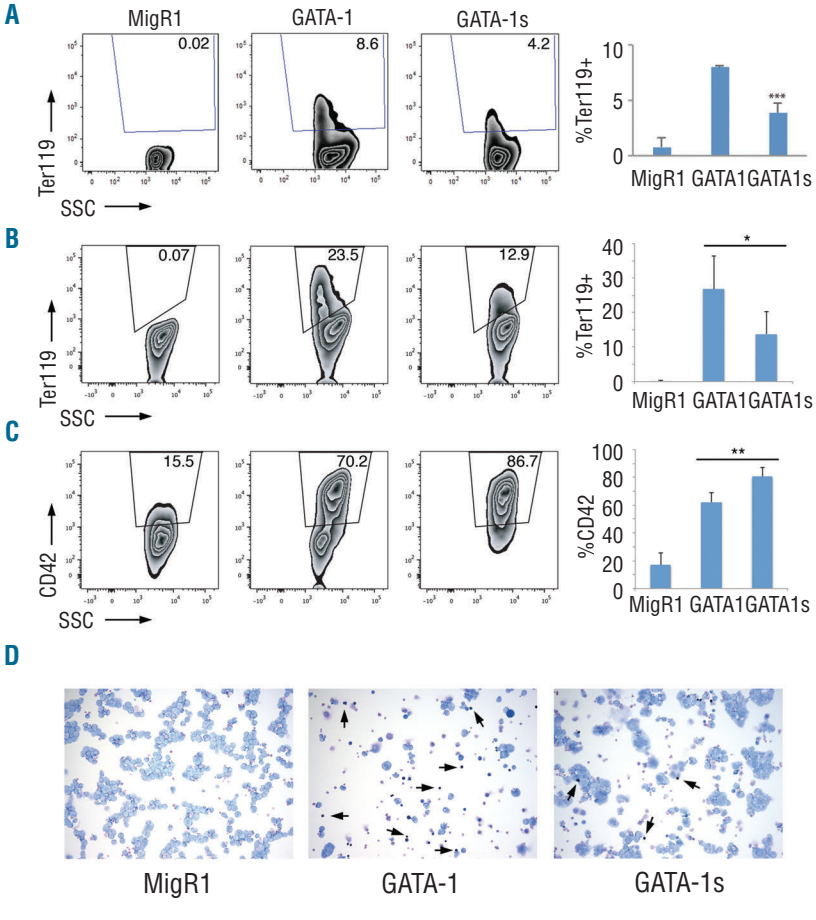

Figure 2. GATA1s is less efficient at promoting erythroid differentiation. (**A**) Flow cytometry analysis of Ter119 expression for G1ME cells cultured in THPO for 72-h post transduction with MigR1 control, GATA1, or GATA1s. (Left) Representative flow plots. (Right) Bar graph showing mean±SD. (**B** and **C**) Flow cytometry analysis of Ter119 (**B**) and CD42 (**C**) expression of G1ME cells cultured in EPO, THPO and SCF 72 h post transduction with MigR1 control, GATA1, or GATA1s viruses. (Left) Representative flow plots. (Right) bar graph depicting the means ± SD. For (**A-C**) n=4; *P<0.05, **P<0.01, ***P<0.001. (**D**) Benzidine stained cytospins of G1ME cultures 72 h post transduction. Arrows point to a subset of benzidine positive erythroid cells. The slides were viewed with a Leica DM4000B microscope fitted with a 20X Leica HCX PL Fluorotar objective. Images were acquired with Leica DFC320 camera and Leica LAS v.4.4 software.

Identification of GATA1s-deficient GATA1 binding sites

Since GATA1s appeared to be deficient in occupancy at many GATA1 binding sites, we sought to identify a set of high confidence GATA1s-deficient sites for systematic analysis. To do this, we reduced the stringency of the peak call for GATA1s from the default setting of 30 down to 10. This resulted in a set of 16,195 GATA1s peaks. We then overlapped this expanded peak set with the original 2728 GATA1 peaks and found that 743 GATA1 peaks still did not have an overlapping GATA1s peak. We refer to these peaks as GATA1s-deficient sites (Figure 4A). Next, we sought to confirm the reduced occupancy of GATA1s at the GATA1s-deficient sites by ChIP-qPCR. We selected a panel of the GATA1s-deficient sites with strong GATA1 occupancy and assayed the occupancy of both GATA1 and GATA1s at these sites in ChIP experiments using both the HA-tag antibody and a GATA1-specific antibody. Using the HA-tag antibody, we found that every tested site was bound at least 1-fold greater by GATA1 than GATA1s (Figure 4B). Using the GATA1 antibody, 19 of 26 (73%) sites were validated by the same standard (Figure 4C). It is clear that the GATA1 antibody yielded higher levels of enrichment for GATA1s than the HA-tag antibody, perhaps indicating some differential sensitivity of the antibodies for the GATA1s molecule. While ChIPs with the GATA1 antibody revealed that many of the GATA1s-deficient sites are occupied at a low level by GATA1s, these data still confirm that there exists a set of sites where GATA1s occupancy is decreased relative to GATA1. Because these quantitative differences in occupancy may be biologically important, we characterized this set of GATA1s-deficient sites in more detail.

We analyzed the location of the GATA1s-deficient sites relative to the nearest transcription start site and found that the sites were distributed similarly to the full GATA1 dataset (Figure 4D). We then analyzed the enrichment of DNA sequence motifs in the GATA1s-deficient sites and found enrichment for the WGATAR motif, an ETS motif, and a CACCC-box motif (Figure 4E). Since ETS proteins and several CACCC-box binding proteins, including KLF1, are critical GATA1 co-regulators, the enrichment of their binding motifs likely indicates that these sites are functional GATA sites. We then searched for motifs that are enriched in the GATA1s-deficient sites relative to the whole set of GATA1 binding sites and found that the GATA and ETS motifs were not enriched, but the CACCC-box was enriched (P=1E-17). This suggests that GATA1s is deficient for occupancy at sites where GATA1 interacts with CACCC-box binding proteins, such as KLF1.

Gene expression analysis reveals that GATA1s fails to properly induce the erythroid differentiation program in G1ME cells

In order to determine the effect of altered chromatin occupancy by GATA1s on gene expression, we completed gene expression profiling of mRNA isolated from control, GATA1-, and GATA1s-expressing G1ME cells 72 h post transduction. Consistent with the differentiation of the cells, 3139 genes were differentially expressed at least 1.5-fold by GATA1 relative to MigR1 control-transduced cells. Similarly, 3079 genes were differentially expressed by GATA1s. Direct comparison of the GATA1 and GATA1s gene expression profiles revealed 847 genes that are differ-

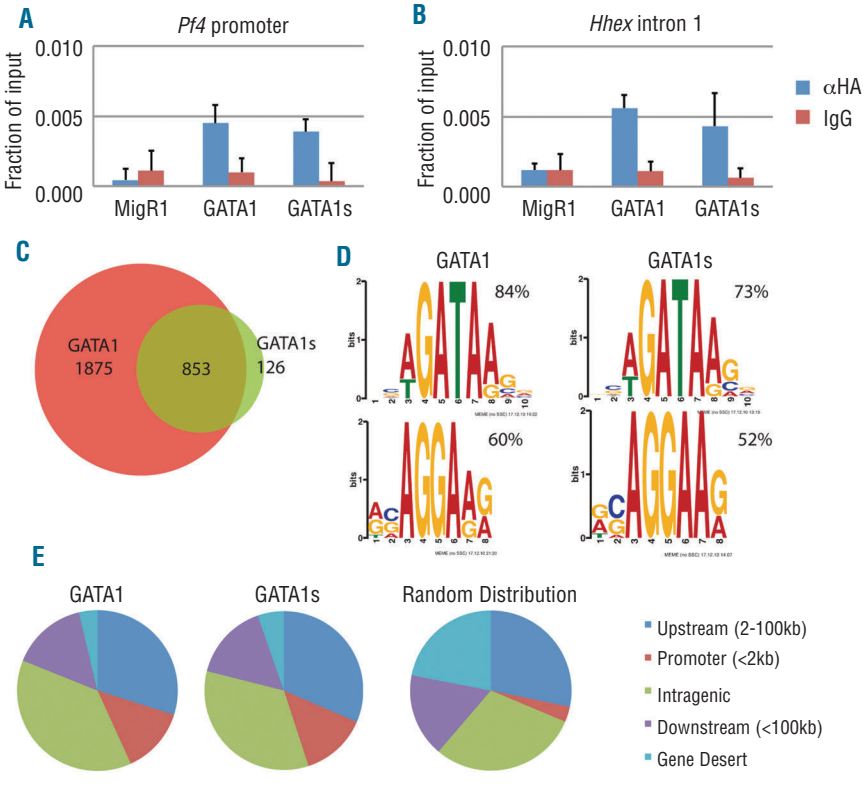

Figure 3. ChIP-Seq results for GATA1 and GATA1s in G1ME cells. (A) ChIP-qPCR using anti-HA tag antibody for the promoter region of the *Pf4* gene in GATA1 and GATA1s-expressing G1ME cells. (B) ChIP-qPCR using anti-HA tag antibody for a known GATA1 binding site in intron 1 of *Hhex* in GATA1 and GATA1s-expressing G1ME cells. (C) Base-wise overlap of GATA1 and GATA1s ChIP-Seq peaks. (D) *De novo* motif finding on 200bp sequence surrounding the centers of the GATA1 and GATA1s peaks. The percentage of peaks that contain the motif is indicated. (E) Depiction of the location of the GATA1 and GATA1s peaks relative to the nearest transcription start site.

entially expressed (Figure 5A). There was no trend toward GATA1-activated or repressed genes being differentially regulated.

We examined the top 10 most differentially regulated genes in either direction and found that many of the genes down-regulated in GATA1s relative to GATA1 were genes that are specific to the erythroid lineage (Figure 5B). Of note, changes were detected in such key factors such as *Hbax*, an embryonic expressed hemoglobin alpha subunit, *Alas2*, an enzyme in the heme biosynthetic pathway, *Tmod1*, a tropomodulin that is required for the integrity of the erythrocyte membrane, *Eraf*, alpha hemoglobin stabilizing protein, and *Slc4a1*, the erythrocyte membrane protein Band3. These changes suggest that GATA1 induces the erythroid gene expression program in G1ME cells cultured in THPO, but GATA1s fails to do so to the same extent. To verify this observation, we determined the enrichment of erythroid-specific genes in the GATA1 and GATA1s gene expression data using Gene Set Enrichment Analysis (GSEA) for a set of genes that are differentially expressed upon erythroid differentiation of G1-ER4 cells (Figure 5C).[21] The GATA1 dataset was significantly enriched for these genes, while the GATA1s dataset was not. Thus, we conclude that GATA1s is much less efficient at inducing the erythroid gene expression program in this cell context as compared to full length GATA1. Of note, we did not find differential enrichment of gene sets for megakaryocytic differentiation [P=0.53, normalized enrichment score (NES)=-0.99], cell cycle (P=0.25, NES=-0.50), or apoptosis (P=0.25, NES=-0.98), indicating that these pathways are not divergently regulated by GATA1s in these cells (*data not shown*).

In order to determine the role of deficient chromatin occupancy by GATA1s in the differential expression of target genes, we overlapped the 752 genes at GATA1s-deficient binding sites with the 847 differentially expressed genes, and found that only 42 are differentially bound and expressed (Figure 6A). Despite this low number, we saw a significant enrichment of differentially expressed genes for differentially bound genes (χ^2=24.4,

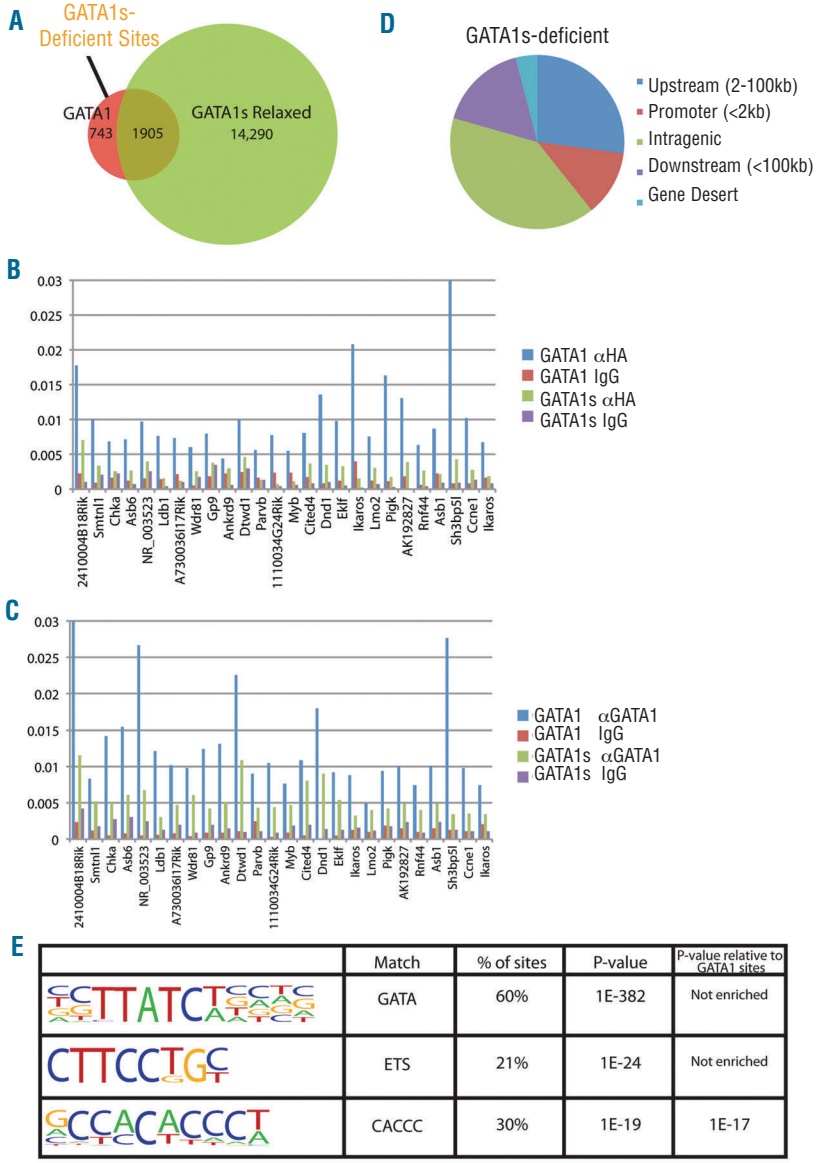

Figure 4. GATA1s is deficient for occupancy at some GATA1 binding sites. (A) Overlap of the original set of GATA1 peaks with an expanded set of GATA1s peaks yields a set of high-confidence GATA1s-deficient binding sites. (B) ChIP-qPCR with an anti-HA antibody to validate reduced chromatin occupancy by GATA1s relative to GATA1 at a panel of GATA1s-deficient sites. (C) ChIP-qPCR with an anti-GATA1 antibody to validate reduced chromatin occupancy by GATA1s relative to GATA1 at a panel of GATA1s-deficient sites. (D) Pie chart of the location of each GATA1s-deficient peak relative to the nearest transcription start site. (E) The top three hits from *de novo* motif analysis on the GATA1s-deficient GATA1 binding sites. *P* values indicate the enrichment of each motif relative to random chance and relative to the whole set of GATA1 peaks.

P=7.87E-7). We classified the genes as up- or down-regulated and found that 31 of the 42 genes are down-regulated in GATA1s cells relative to GATA1 (Table 1). This indicates that the failure of GATA1s to occupy a binding site most often coincides with failure to activate the gene. We then used GSEA to determine the enrichment of genes differentially expressed by GATA1 or GATA1s in the genes that have a GATA1s-deficient binding site (Figure 6B). We found a significant enrichment of the GATA1 gene expression profile, but not that of GATA1s, indicating that GATA1s is deficient for the expression of the genes that it fails to bind, as expected. Importantly, we did not identify significant differences in the expression of genes that are bound by both genes (Figure 6C). We examined the top 10 most-differentially expressed genes from those with GATA1s-deficient binding sites and noticed that many of them were erythroid-specific genes (Figure 6D). Closer examination of the GATA1 occupancy at these genes revealed that GATA1s is deficient for occupancy at the first introns of *Slac4a1* and *Alas2* and the promoter of *Klf1*, and that it fails to activate the expression of these genes to the same extent as GATA1 (Figure 6E). We confirmed that GATA1s is deficient for occupancy at these sites in G1ME cells cultured in both EPO and THPO by ChIP-qPCR using the GATA1-specific antibody (Figure 6F).

Discussion

In this report, we show that GATA1s is deficient for erythroid specific gene expression activity, but proficient for megakaryocyte specific gene expression in cells with both erythroid and megakaryocytic differentiation potential. This finding is consistent with recent reports of anemia in patients bearing germ-line GATA1s mutations.[29,30] Reduced erythroid gene expression in uncommitted progenitor cells would likely result in reduced commitment to the ery-

A

B

Gene Name	Fold change	Direction (G1s:G1)
Hba-x	571.10486	down
Car1	345.50525	down
Tnni3	54.84258	down
Alas2	41.771973	down
Tmod1	37.827873	down
Eraf	28.956932	down
Slc4a1	24.076063	down
Plxdc1	22.447338	down
Crlf1	19.302702	down
Tmem184a	16.362177	down
Slc1a2	9.89985	up
Hrc	10.300123	up
Rpgrip1	10.642997	up
Ccl19	10.951774	up
Fbp3	10.962054	up
Nenf	11.854194	up
C230016C14Rik	12.475878	up
EG433229	14.116186	up
Dnmt3l	44.99297	up
Ccl9	50.55146	up

C

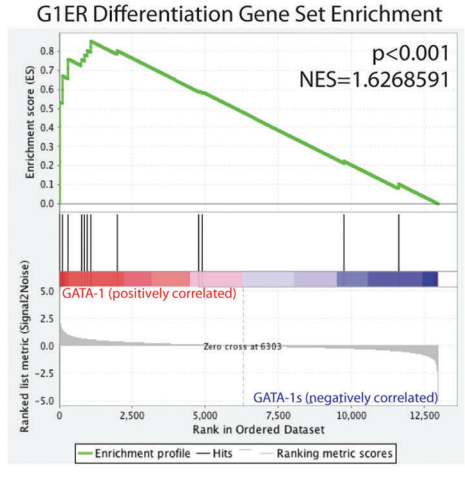

Figure 5. Global gene expression analysis reveals differential gene expression in GATA1s-expressing G1ME cells relative to GATA1. (A) Heatmap of gene expression relative to MigR1 control for 847 genes differentially expressed in GATA1 and GATA1s expressing cells. (B) A list of the top 10 most up- and down-regulated genes in GATA1s-expressing G1ME cells relative to GATA1-expressing cells. (C) Gene set enrichment plot for a set of genes differentially expressed upon erythroid differentiation of G1-ER4 cells on gene expression data from GATA1 and GATA1s-expressing G1ME cells.

throid lineage leading to profound anemia.

Using both ChIP-Seq and gene expression profiling, we found that GATA1s was both deficient for chromatin occupancy and the activation of gene expression at many erythroid specific genes. The reduced occupancy at erythroid genes was unexpected since GATA1s maintains both zinc finger DNA binding domains and its interaction with its essential co-factor FOG1.[23] Since GATA1s occupies and activates megakaryocyte-specific genes with normal efficiency, it is likely that GATA1s is deficient for erythroid specific co-factor interactions that stabilize its chromatin occupancy, or perhaps it fails to activate the expression of factors that precede its occupancy at erythroid sites. KLF1 is an enticing candidate for either of these potential mechanisms since it is known to interact with GATA1 and co-occupy gene regulatory regions, its target sequence was enriched in the GATA1s-deficient binding sites, and because GATA1s fails to activate its expression. However, the interaction with KLF1 has been mapped to the C-terminal zinc finger and thus is expected to be maintained in GATA1s.[13,44]

Very few GATA1 protein interactions have been mapped to the N-terminal domain. One study indicated that RUNX1 binds this domain, but this interaction is thought to be required for megakaryocyte specification, but not erythroid.[45] Furthermore, a different study indicated the RUNX-GATA1 interaction occurs through the zinc finger domains and is maintained by the GATA1s isoform.[46] Another study demonstrated an interaction with pRB that is ablated by the GATA1s mutation and further revealed that the protein-protein interaction is required for cell cycle arrest during terminal erythroid differentiation.[47] Loss of the Rb interaction is also thought to contribute to leukemic transformation in DS-AMKL through increased E2F activity.[48]

It is possible that there are multiple factors that contribute to the erythroid specific deficiency caused by GATA1s mutations. A promising candidate is LMO2, a component of the SCL pentameric complex, which is known to co-operate with GATA1 in gene activation but not gene repression.[49] We found that the expression of LMO2 was decreased in GATA1s-expressing G1ME cells relative to those expressing GATA1. Given, however, that the SCL complex and GATA-1 also co-regulate megakaryocyte-specific genes, it is unlikely that a deficiency in LMO2 causes an erythroid-specific defect. Future identification of proteins that bind the GATA1 N-terminal domain is necessary to identify these important interactions.

Previous studies on the transcriptional activity of GATA1s were performed with either committed erythroid cells or megakaryocytes, and thus the defect in lineage commitment that we report has not previously been shown. For example, Weiss and colleagues reported that

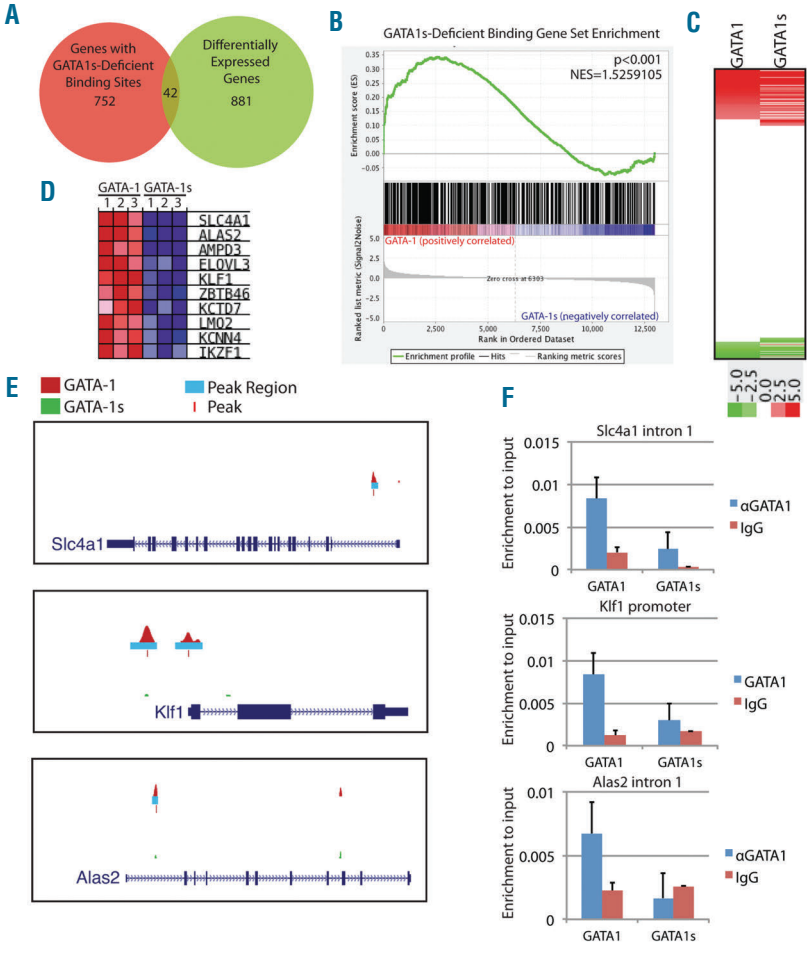

Figure 6. GATA1s is deficient for occupancy and activation of several erythroid-specific genes. (A) Venn diagram of the overlap of genes at GATA1s-deficient binding sites and genes differentially expressed by GATA1s relative to GATA1. (B) GSEA for the set of genes at GATA1s deficient binding sites on the gene expression data from GATA1 and GATA1s-expressing cells. (C) A heatmap depicting the expression of genes bound by both GATA1 and GATA1s. (D) The top 10 most significantly enriched genes from the GSEA analysis in (B) are depicted in a heatmap. (E) UCSC genome browser depictions of ChIP Seq data at GATA1s-deficient binding sites. GATA1 and GATA1s wig profiles are displayed in red and green, respectively. The light blue bars indicate enriched regions and the red dashes indicate the peak centers. (F) ChIP-qPCR for GATA1 and GATA1s in G1ME cells using anti-GATA1 antibody. Bar graphs represent average±SD of two independent experiments.

expression of a GATA1 mutant lacking the N-terminal activation domain was able to promote differentiation of the G1E proerythroblast cell line.[21] Given this difference with our work, it is likely that the GATA1s defect lies at the lineage specification stage: once past that step, GATA1s appears to be able to drive terminal differentiation. It is notable that the fetal livers of Gata1s knock-in mice display a prominent expansion of megakaryocytes that is accompanied by deficiency in erythroid cells.[50] This latter observation is consistent with our findings that GATA1s is defective in erythroid specification, gene expression, and chromatin occupancy.

Finally, although we observed a selective defect in chromatin occupancy of GATA1s in erythroid cells, Klusmann and colleagues recently reported that GATA1s showed impaired occupancy at the Myc promoter in eosinophils.[51] This defect was accompanied by impaired Myc gene repression and an expansion in eosinophils. Thus, the regulation of GATA1 chromatin binding and transcriptional activity is controlled by the N-terminus in multiple hematopoietic cells.

Acknowledgments

The authors thank Lou Dore for data analysis and helpful discussion, and Jennifer Busuttil and Ari Melnick of Cornell University for processing ChIP-Seq samples.

Funding

This research was supported by a grant from the NIH (DK101329) and the Samuel Waxman Cancer Research Foundation.

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Wu W, Cheng Y, Keller CA, et al. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. Genome Res. 2011;21(10):1659-1671.
2. Cheng Y, Wu W, Ashok Kumar S, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res. 2009; 19(12):2172-2184.
3. Stachura DL, Chou ST, Weiss MJ. Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. Blood. 2006;107(1):87-97.
4. 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(5):597-609.
5. Chlon TM, Dore LC, Crispino JD. Cofactor-Mediated Restriction of GATA-1 Chromatin Occupancy Coordinates Lineage-Specific Gene Expression. Mol Cell. 2012;47(4):608-621.
6. Chlon TM, Crispino JD. Combinatorial regulation of tissue specification by GATA and FOG factors. Development. 2012; 139(21):3905-3916.
7. Dore LC, Crispino JD. Transcription factor networks in erythroid cell and megakaryocyte development. Blood. 2011; 118(2):231-239.
8. Lowry JA, Mackay JP. GATA-1: One protein, many partners. Int J Biochem Cell Biol. 2006;38(1):6-11.
9. Pang L, Xue H-H, Szalai G, et al. Maturation stage specific regulation of megakaryopoiesis by pointed-domain Ets proteins. Blood. 2006;108(7):2198-2206.
10. Eisbacher M, Holmes ML, Newton A, et al. Protein-Protein Interaction between Fli-1 and GATA-1 Mediates Synergistic Expression of Megakaryocyte-Specific Genes through Cooperative DNA Binding. Mol Cell Biol. 2003;23(10):3427-3441.
11. Wang X, Crispino JD, Letting DL, Nakazawa M, Poncz M, Blobel GA. Control of megakaryocyte-specific gene expression by GATA-1 and FOG-1: role of Ets transcription factors. EMBO J. 2002; 21(19):5225-5234.
12. Crossley M, Tsang AP, Bieker JJ, Orkin SH. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. J Biol Chem. 1994;269(22):15440-15444.
13. Merika M, Orkin SH. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. Mol Cell Biol. 1995;15(5):2437-2447.
14. Chen X, Reitman M, Bieker JJ. Chromatin Structure and Transcriptional Control Elements of the Erythroid Krüppel-like Factor (EKLF) Gene. J Biol Chem. 1998; 273(39):25031-25040.
15. Anderson KP, Crable SC, Lingrel JB. The GATA-E box-GATA motif in the EKLF promoter is required for in vivo expression. Blood. 2000;95(5):1652-1655.
16. 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(2):219-228.
17. Tallack MR, Whitington T, Shan Yuen W, Wainwright EN, Keys JR, Gardiner BB, et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res. 2010;20(8):1052-1063.
18. 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(17):e139-e148.
19. Calligaris R, Bottardi S, Cogoi S, Apezteguia I, Santoro C. Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. Proc Natl Acad Sci USA. 1995;92(25):11598-11602.
20. Martin DI, Orkin SH. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes Dev. 1990;4(11):1886-1898.
21. Weiss MJ, Yu C, Orkin SH. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. Mol Cell Biol. 1997;17(3):1642-1651.
22. Muntean AG, Crispino JD. Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. Blood. 2005;106(4):1223-1231.
23. Wechsler J, Greene M, McDevitt MA, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat Genet. 2002;32(1):148-152.
24. Mundschau G, Gurbuxani S, Gamis AS, Greene ME, Arceci RJ, Crispino JD. Mutagenesis of GATA1 is an initiating event in Down syndrome leukemogenesis. Blood. 2003;101(11):4298-4300.
25. Khan I, Malinge S, Crispino J. Myeloid leukemia in Down syndrome. Crit Rev Oncog. 2011;16(1-2):25-36.
26. Roberts I, Alford K, Hall G, et al. GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia. Blood. 2013;122(24):3908-3917.
27. Nikolaev SI, Santoni F, Vannier A, et al. Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down syndrome. Blood. 2013;122(4):554-561.
28. Yoshida K, Toki T, Okuno Y, et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nat Genet. 2013;45(11):1293-1299.
29. Hollanda LM, Lima CSP, Cunha AF, et al. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. Nat Genet. 2006;38(7):807-812.
30. Sankaran VG, Ghazvinian R, Do R, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J Clin Invest. 2012;122(7):2439-2443.
31. Parrella S, Aspesi A, Quarello P, et al. Loss of GATA-1 full length as a cause of Diamond-Blackfan anemia phenotype. Pediatr Blood Cancer. 2014;61(7):1319-1321.
32. Huang Z, Richmond TD, Muntean AG, Barber DL, Weiss MJ, Crispino JD. STAT1 promotes megakaryopoiesis downstream of GATA-1 in mice. J Clin Invest. 2007; 117(12):3890-3899.

33. Forsberg EC, Downs KM, Bresnick EH. Direct interaction of NF-E2 with hypersensitive site 2 of the β-globin locus control region in living cells. Blood. 2000;96(1):334-339.
34. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science. 2007; 316(5830):1497-1502.
35. Valouev A, Johnson DS, Sundquist A, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods. 2008;5(9):829-834.
36. Blahnik KR, Dou L, O'Geen H, et al. Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. Nucleic Acids Res. 2010;38(3):e13.
37. Dore LC, Chlon TM, Brown CD, White KP, Crispino JD. Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. Blood. 2012;119(16):3724-3733.
38. Bailey TL, Boden M, Buske FA, et al. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37(suppl 2):W202-W8.
39. 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(4):576-589.
40. de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004;20(9):1453-1454.
41. Saldanha AJ. Java Treeview -- extensible visualization of microarray data. Bioinformatics. 2004;20(17):3246-3248.
42. Shimizu R, Takahashi S, Ohneda K, Engel JD, Yamamoto M. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. EMBO J. 2001;20(18):5250-5260.
43. Dore LC, Chlon TM, Brown CD, White KP, Crispino JD. Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. Blood. 2012;119(16):3724-3733.
44. Gregory R, Taxman D, Seshasayee D, Kensinger M, Bieker J, Wojchowski D. Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. Blood. 1996; 87(5):1793-1801.
45. Elagib KE, Racke FK, Mogass M, Khetawat R, Delehanty LL, Goldfarb AN. RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. Blood. 2003;101(11):4333-4341.
46. Xu G, Kanezaki R, Toki T, et al. Physical association of the patient-specific GATA1 mutants with RUNX1 in acute megakaryoblastic leukemia accompanying Down syndrome. Leukemia. 2006;20(6):1002-1008.
47. Kadri Z, Shimizu R, Ohneda O, et al. Direct Binding of pRb/E2F-2 to GATA-1 Regulates Maturation and Terminal Cell Division during Erythropoiesis. PLoS Biol. 2009; 7(6):e1000123.
48. Klusmann J-H, Godinho FJ, Heitmann K, et al. Developmental stage-specific interplay of GATA1 and IGF signaling in fetal megakaryopoiesis and leukemogenesis. Genes Dev. 2010;24(15):1659-1672.
49. Tripic T, Deng W, Cheng Y, et al. SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. Blood. 2009;113(10):2191-2201.
50. Li Z, Godinho FJ, Klusmann J-H, Garriga-Canut M, Yu C, Orkin SH. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. Nat Genet. 2005;37(6):613-619.
51. Maroz A, Stachorski L, Emmrich S, et al. GATA1s induces hyperproliferation of eosinophil precursors in Down syndrome transient leukemia. Leukemia. 2014; 28(6):1259-1270.