JCI The Journal of Clinical Investigation

Pluripotent stem cells reveal erythroid-specific activities of the GATA1 N-terminus

Marta Byrska-Bishop, … , Mitchell J. Weiss, Stella T. Chou

J Clin Invest. 2015;125(3):993-1005. https://doi.org/10.1172/JCI75714.

Research Article **Stem cells**

Germline *GATA1* mutations that result in the production of an amino-truncated protein termed GATA1s (where s indicates short) cause congenital hypoplastic anemia. In patients with trisomy 21, similar somatic GATA1s-producing mutations promote transient myeloproliferative disease and acute megakaryoblastic leukemia. Here, we demonstrate that induced pluripotent stem cells (iPSCs) from patients with GATA1-truncating mutations exhibit impaired erythroid potential, but enhanced megakaryopoiesis and myelopoiesis, recapitulating the major phenotypes of the associated diseases. Similarly, in developmentally arrested GATA1-deficient murine megakaryocyte-erythroid progenitors derived from murine embryonic stem cells (ESCs), expression of GATA1s promoted megakaryopoiesis, but not erythropoiesis. Transcriptome analysis revealed a selective deficiency in the ability of GATA1s to activate erythroid-specific genes within populations of hematopoietic progenitors. Although its DNA-binding domain was intact, chromatin immunoprecipitation studies showed that GATA1s binding at specific erythroid regulatory regions was impaired, while binding at many nonerythroid sites, including megakaryocytic and myeloid target genes, was normal. Together, these observations indicate that lineage-specific GATA1 cofactor associations are essential for normal chromatin occupancy and provide mechanistic insights into how *GATA1s* mutations cause human disease. More broadly, our studies underscore the value of ESCs and iPSCs to recapitulate and study disease phenotypes.

Find the latest version:

https://jci.me/75714/pdf

Pluripotent stem cells reveal erythroid-specific activities of the GATA1 N-terminus

Marta Byrska-Bishop,[1] Daniel VanDorn,[2] Amy E. Campbell,[2] Marisol Betensky,[2] Philip R. Arca,[2] Yu Yao,[3] Paul Gadue,[4] Fernando F. Costa,[5] Richard L. Nemiroff,[6] Gerd A. Blobel,[2] Deborah L. French,[4] Ross C. Hardison,[1] Mitchell J. Weiss,[3] and Stella T. Chou[2]

[1]Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, USA. [2]Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [3]Department of Hematology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [4]Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [5]Department of Internal Medicine, School of Medicine, University of Campinas, Campinas, Brazil. [6]Department of Obstetrics and Gynecology, The University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

Germline *GATA1* mutations that result in the production of an amino-truncated protein termed GATA1s (where s indicates short) cause congenital hypoplastic anemia. In patients with trisomy 21, similar somatic GATA1s-producing mutations promote transient myeloproliferative disease and acute megakaryoblastic leukemia. Here, we demonstrate that induced pluripotent stem cells (iPSCs) from patients with GATA1-truncating mutations exhibit impaired erythroid potential, but enhanced megakaryopoiesis and myelopoiesis, recapitulating the major phenotypes of the associated diseases. Similarly, in developmentally arrested GATA1-deficient murine megakaryocyte-erythroid progenitors derived from murine embryonic stem cells (ESCs), expression of GATA1s promoted megakaryopoiesis, but not erythropoiesis. Transcriptome analysis revealed a selective deficiency in the ability of GATA1s to activate erythroid-specific genes within populations of hematopoietic progenitors. Although its DNA-binding domain was intact, chromatin immunoprecipitation studies showed that GATA1s binding at specific erythroid regulatory regions was impaired, while binding at many nonerythroid sites, including megakaryocytic and myeloid target genes, was normal. Together, these observations indicate that lineage-specific GATA1 cofactor associations are essential for normal chromatin occupancy and provide mechanistic insights into how *GATA1s* mutations cause human disease. More broadly, our studies underscore the value of ESCs and iPSCs to recapitulate and study disease phenotypes.

Introduction

GATA1 is an essential erythro-megakaryocytic transcription factor. In normal humans, alternative splicing or translation initiation site usage produces 2 forms of GATA1 protein: full-length (GATA1fl) and a truncated isoform lacking the first 83 amino acids (GATA1s) (1, 2). Germline *GATA1* mutations that favor GATA1s expression over full-length GATA1 (hereafter referred to as "*GATA1s* mutations") were identified in a pedigree with congenital hypoplastic anemia and neutropenia (2) and in patients with Diamond Blackfan anemia (DBA), a rare congenital anemia with erythroid hypoplasia (3, 4). In individuals with germline trisomy 21 (T21, Down syndrome [DS]), somatic *GATA1s* mutations in fetal hematopoietic progenitors promote 2 clonal disorders, transient myeloproliferative disease (TMD) and acute megakaryoblastic leukemia (AMKL) (5, 6). Notably, euploid patients with germline *GATA1s* mutations are not predisposed to leukemia.

The GATA1 N-terminus, which is absent in GATA1s, was originally named the N-terminal activation domain (NAD) by virtue of its ability to activate transcription in heterologous cells (1, 7). However, the GATA1 N-terminus is partially dispensable for hematopoiesis in several murine models (8–10). *Gata1*$^{-}$ embryos all die of profound anemia in midgestation (E10.5) (11). In contrast, mice with germline *Gata1s* mutations exhibit moderate-to-severe anemia and enhanced megakaryopoiesis between E9.5 and E14.5, but some survive to birth, after which hematopoiesis normalizes postnatally (12, 13). Overall, studies of human patients and genetically manipulated mice demonstrate that the functions of the GATA1 N-terminus and the consequences of its loss depend on gene dosage, cell context, developmental stage, and species (2, 3, 9, 12–14).

Defining how the GATA1 N-terminus (or lack thereof) affects hematopoiesis separately or in combination with T21 has been challenging, partly due to lack of ideal experimental systems. We investigated the functions of the GATA1 N-terminus by analyzing induced pluripotent stem cells (iPSCs) from patients with GATA1s mutations and *Gata1* gene–disrupted murine embryonic stem cells (ESCs). Both ESCs and iPSCs self renew in culture and generate multiple hematopoietic lineages through directed in vitro differentiation (15, 16). Our findings provide 2 major mechanistic insights into functions of the GATA1 N-terminus and pathophysiologies of human diseases associated with its loss. First, *GATA1s* mutations impair erythropoiesis, with hematopoietic output biased toward myeloid and megakaryocytic cells. Second, GATA1s binding is specifically

Authorship note: Marta Byrska-Bishop and Daniel VanDorn contributed equally to this work.
Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: February 17, 2014; **Accepted:** December 15, 2014.
Reference information: *J Clin Invest*. 2015;125(3):993–1005. doi:10.1172/JCI75714.

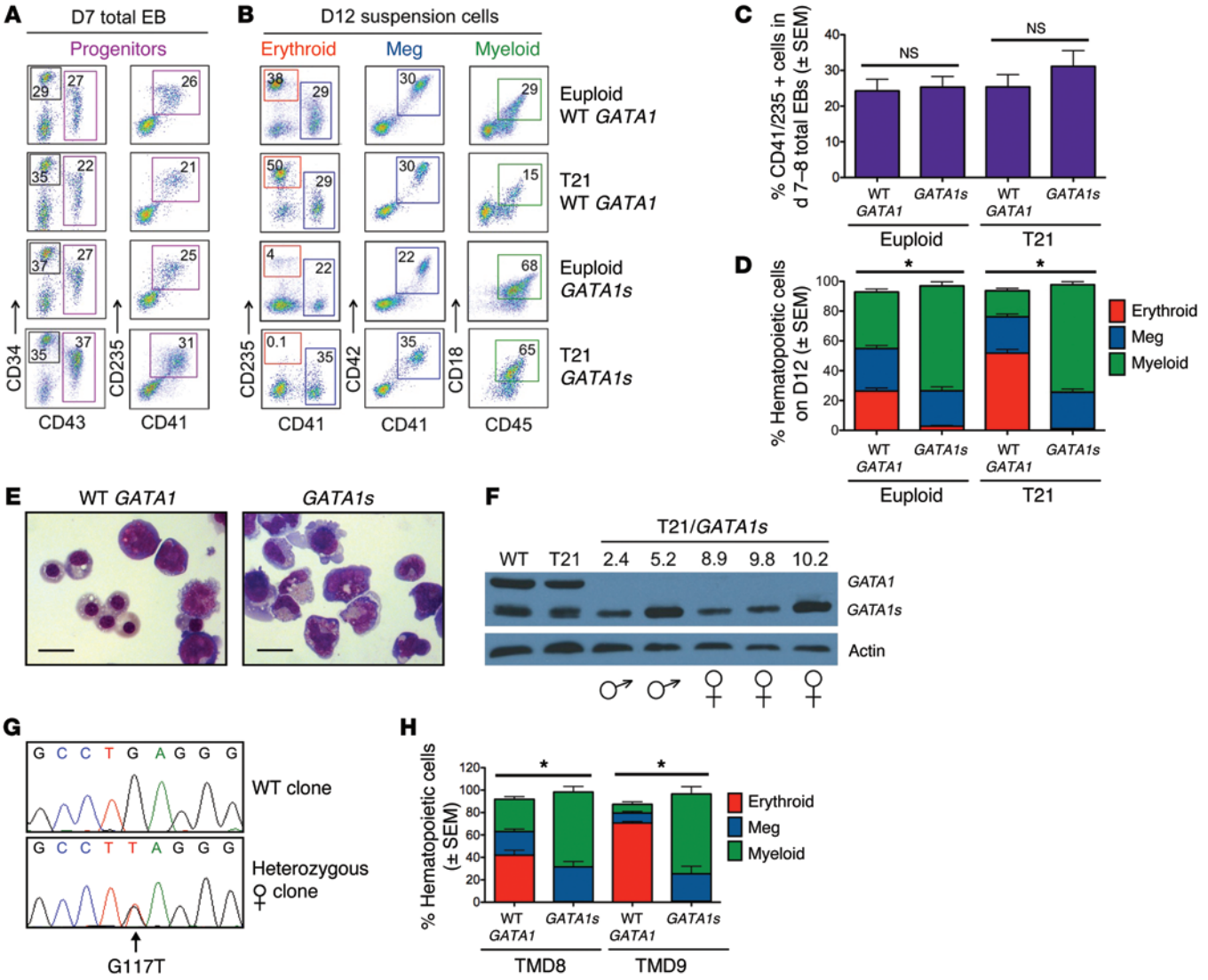

Figure 1. *GATA1s* mutations inhibit erythropoiesis from patient-derived iPSCs. (**A**) Flow-cytometry analysis of CD34$^{+/-}$CD43^{+}CD41^{+}CD235^{+} progenitors within total EB cultures on day 7 of hematopoietic differentiation and (**B**) suspension cells released from EBs on day 12 showing mature hematopoietic lineages: erythroid (CD41^{-}CD235^{+}), megakaryocytic (Meg, CD41^{+}CD42^{+}), and myeloid (CD45^{+}CD18^{+}). Numbers denote percentage of total cells in the indicated gate. (**C**) Frequency of CD43^{+}CD41^{+}CD235^{+} progenitor cells in EB cultures on days 7 and 8 of hematopoietic differentiation (n = 6; 17 independent experiments for euploid and T21 groups, respectively). (**D**) Summary of distribution of lineage-committed cells in EB suspension cultures on differentiation day 12 (n = 12; 20 independent assays for euploid and T21 groups, respectively). (**E**) Hematopoietic cell morphology on day 20 of differentiation cultures of isogenic WT *GATA1* or *GATA1s* iPSCs. Scale bars: 50 μm. (**F**) Western blot of iPSC-derived hematopoietic cells. (**G**) DNA sequence analysis showing WT *GATA1* and a heterozygous exon 2 mutation in 2 different iPSC clones from a female with DS and TMD. (**H**) Isogenic lines from 2 different TMD patients were analyzed. Percentages of mature lineages from day-12 EB suspension cultures, as in **D**. (n = 6; 4 independent assays for TMD8 and TMD9, respectively). *P < 0.005 for myeloid and erythroid lineages (2-tailed Student's t test).

impaired at erythroid target genes, implicating the N-terminus as a selective chromatin occupancy factor. More generally, our studies illustrate the utility of pluripotent stem cells as model systems for investigating human diseases.

Results

Generation of iPSCs. We reprogrammed peripheral blood mononuclear cells from 5 DS patients with TMD caused by somatic *GATA1s* mutations and 1 euploid patient with a similar germline *GATA1* mutation (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI75714DS1). Additional control iPSC lines were produced from fibroblasts or primary stromal cells of T21 or euploid subjects with WT *GATA1*. Most cells were reprogrammed using a polycistronic lentiviral vector encoding *OCT4*, *SOX2*, *KLF4*, and *MYC* regulated by a doxycycline-inducible promoter (17, 18). Early in this study, some lines were generated using individual retroviruses, each encoding 1 reprogramming factor. We obtained similar results for each genotype created using different reprogramming systems. In DS-TMD, premalignant blasts carrying the *GATA1s* mutation circulate with WT *GATA1* blood cells. Reprogramming blood mononuclear cells from such patients generated "isogenic" groups of *GATA1s* and WT *GATA1* iPSC clones (n = 2 different patients, Supplemental Table 1). All iPSC clones fulfilled standard quality control criteria

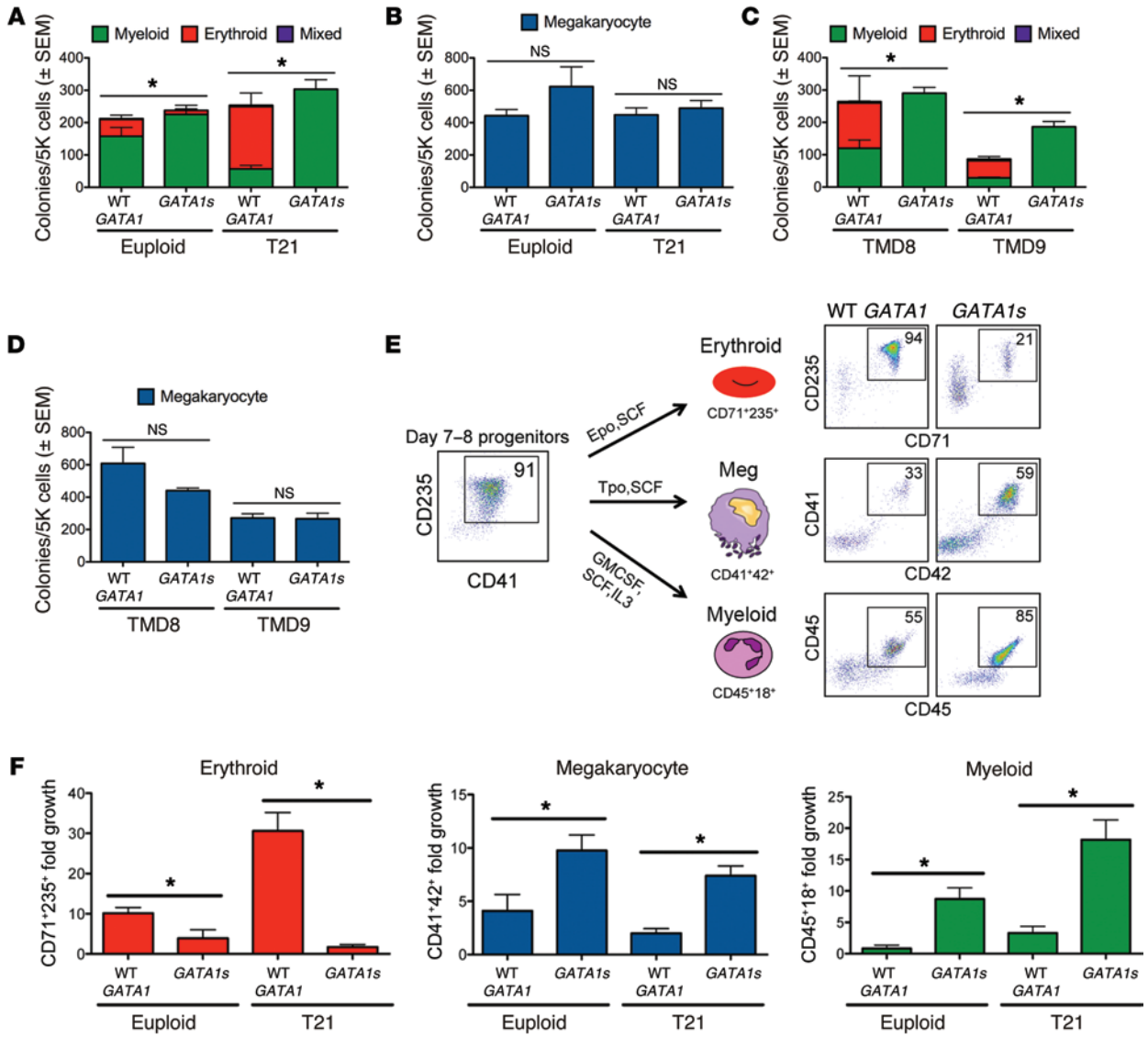

Figure 2. GATA1s mutations impair erythropoiesis and enhance the production of myeloid cells and megakaryocytes in iPSC-derived CD43$^+$CD41$^+$CD235$^+$ progenitors. (**A**) Methylcellulose colony assays containing SCF, IL-3, EPO, GM-CSF, and (**B**) colony-forming megakaryocyte (CFU-Mk) assays with TPO, IL-3, and IL-6 of CD43$^+$CD41$^+$CD235$^+$ progenitors. Results show mean values ± SEM of all lines performed in triplicate (n = 4 euploid/WT GATA1, 2 euploid/GATA1s, 4 T21/WT GATA1, and 5 T21/GATA1s lines). (**C**) Methylcellulose colony assays and (**D**) CFU-Mk formation from CD43$^+$CD41$^+$CD235$^+$ progenitors from 2 pairs of isogenic iPSC clones with the indicated *GATA1* alleles, performed as in **A** and **B**. (**E**) Schematic and representative flow-cytometry analysis of WT *GATA1* and *GATA1s* CD43$^+$CD41$^+$CD235$^+$ progenitors grown in liquid culture to preferentially support erythroid (CD71$^+$CD235$^+$), megakaryocyte (CD41$^+$CD42$^+$), and myeloid (CD45$^+$CD18$^+$) differentiation. Numbers denote percentage of total cells in the indicated gate. (**F**) Fold expansion of erythroid, megakaryocyte, and myeloid cells after 6 days in lineage-specific liquid cultures of euploid or T21 progenitors expressing WT GATA1 or GATA1s (n = 4 independent assays for each bar with the exception of n = 3 for euploid/*GATA1s* in erythroid culture, n = 5 for T21/WT *GATA1* in erythroid culture, and n = 5 for T21/WT *GATA1* and T21/*GATA1s* in megakaryocyte culture). *$P < 0.05$ (2-tailed Student's t test).

(Supplemental Figure 1). DNA sequencing detected the appropriate *GATA1* mutations and karyotypes in patient-derived iPSCs (Supplemental Table 1).

Loss of the N-terminus of GATA1 perturbs hematopoiesis. We compared the blood-forming potential of euploid and T21 iPSC lines with WT *GATA1* or *GATA1s* mutations. We generated embryoid bodies (EBs) in defined media containing sequential combinations of cytokines to induce primitive streak/mesoderm formation followed by hematopoietic specification and multilineage differentiation (Supplemental Figure 2 and refs. 15, 18). On days 7 to 8 of EB differentiation, a hematopoietic progenitor population (CD43$^+$CD41$^+$CD235$^+$) was present at similar frequencies within EBs of all genotypes (Figure 1, A and C). These progenitors were released into the medium, and by day 12, WT *GATA1* control cultures contained lineage-committed erythroid (CD235$^+$CD41$^-$), megakaryocytic (CD41$^+$CD42$^+$), and myeloid (CD45$^+$CD18$^+$) cells (Figure 1B, top row). T21/WT *GATA1* iPSCs produced about 2-fold increased proportion of erythroblasts compared with euploid iPSCs, as described previously (18). Euploid/*GATA1s* and T21/*GATA1s* progenitors exhibited severely impaired erythropoiesis despite retained capacity for myeloid and megakaryocytic differentiation (Figure 1B, bottom 2 rows, and Figure 1D). May-Grün-

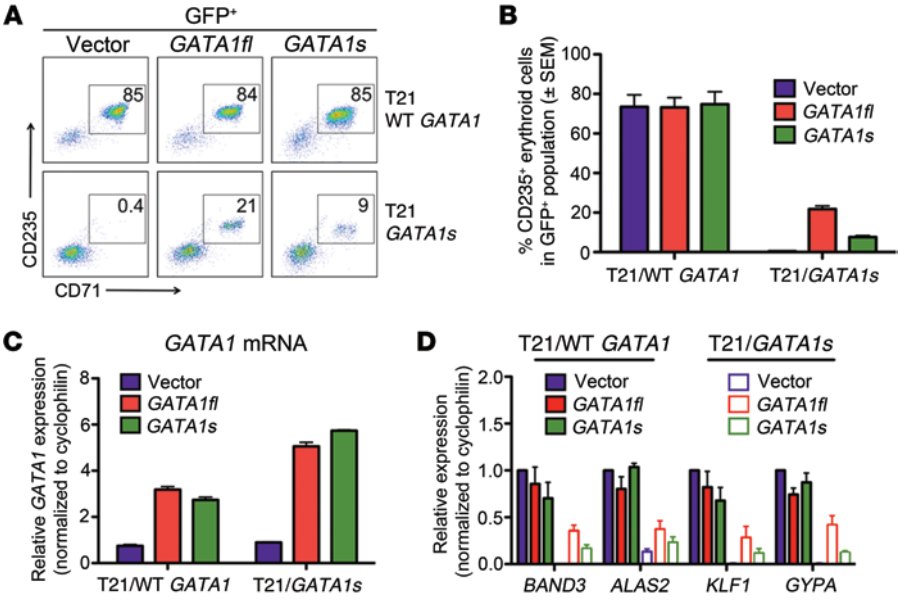

Figure 3. Dose-dependent restoration of erythropoiesis with enforced expression of full-length GATA1 and truncated GATA1s. T21/WT *GATA1* and T21/*GATA1s* iPSC–derived CD43$^+$CD41$^+$CD235$^+$ progenitors were transduced with lentivirus containing vector alone or encoding GATA1fl or GATA1s and cultured with EPO and SCF (n = 4 replicates). (**A**) Representative flow-cytometric analysis after 6 days of culture. Numbers denote percentage of total cells in the indicated gate. (**B**) Average percentage of CD235$^+$ erythroblasts in transduced (GFP$^+$) cells after 6 days of culture. (**C**) PCR showing relative *GATA1* mRNA expression 2 days after lentiviral infection with control vector, *GATA1fl*, or *GATA1s*. The genotype of progenitors is shown on the *x* axis. (**D**) Expression of selected erythroid mRNAs in flow-cytometry–purified CD235$^+$ infected (GFP$^+$) cells.

wald-Giemsa staining revealed numerous late-stage erythroblasts in day-20 EB cultures from WT *GATA1* iPSCs (T21 or euploid), but not from *GATA1s* iPSCs (Figure 1E).

Western blotting of EB-derived hematopoietic cells from euploid and T21 iPSC lines with WT *GATA1* demonstrated predominantly full-length GATA1, but also GATA1s (Figure 1F; see complete unedited blots in the supplemental material). All T21/*GATA1s* iPSC–derived hematopoietic cells expressed GATA1s, but no full-length protein. These findings occurred in male and female iPSCs that were hemizygous and heterozygous for the X chromosome–linked *GATA1s* mutation, respectively. Thus, female T21/*GATA1s* iPSC clones retained the same inactive X chromosome as the somatic cells from which they were derived, consistent with prior reports that X chromosome inactivation is maintained during human iPSC reprogramming (19–21). During the course of this study, serial Western blotting for GATA1 showed that female *GATA1s* iPSC lines did not undergo reactivation of the lyonized X chromosome, which can occur during iPSC passage (19, 22).

Genetic and epigenetic variation between different iPSC lines may alter their developmental potential, including hematopoietic (23, 24). We controlled for this by systematically analyzing multiple iPSC clones from different individuals of the same relevant genotype (Figure 1, C and D, n = 6 *GATA1s* and 12 WT *GATA1*) and isogenic lines from TMD blasts (T21/*GATA1s*) and normal blood cells (T21/WT *GATA1*) of the same DS patients (n = 2 different subjects, Supplemental Table 1 and Figure 1, G and H). The *GATA1s* mutation selectively blocked erythropoiesis in all iPSC clones examined.

We next assessed the hematopoietic colony-forming potential of day-7 to day-9 EB-derived CD43$^+$CD41$^+$CD235$^+$ progenitors from euploid and T21 iPSCs with WT *GATA1* or *GATA1s* genotypes. T21/WT *GATA1* progenitors generated 2- to 3-fold increased proportion of erythroid colonies compared with euploid progenitors (Figure 2A), as previously reported (18). In contrast, erythroid colony-forming potential was severely reduced in *GATA1s* progenitors compared with WT *GATA1* progenitors from iPSCs of different individuals, irrespective of chromosome 21 status (Figure 2A). Megakaryocyte colony-forming frequency and size were normal in *GATA1s* iPSCs (Figure 2B), while myeloid colony-forming potential was increased and myeloid colony size tended to be larger (Figure 2A and Supplemental Figure 3). We obtained similar results comparing 2 sets of isogenic WT *GATA1* and *GATA1s* iPSC lines from individuals with DS-associated TMD (Figure 2, C and D). Together, the EB suspension and methylcellulose colony assays indicate that GATA1s inhibits erythropoiesis and enhances myelopoiesis.

Similar megakaryocyte colony formation between WT *GATA1* and *GATA1s* iPSC–derived progenitors suggested comparable megakaryocyte progenitor frequency between the 2 genotypes. However, collagen-based assays are not a robust method for determining megakaryocyte proliferative capacity. To investigate this property further for all lineages, we cultured EB-derived CD43$^+$CD41$^+$CD235$^+$ progenitors on OP9 stromal cells with cytokine combinations that promote erythroid, megakaryocytic, or myeloid development selectively (Figure 2E). In erythropoietin (EPO) and stem cell factor (SCF), WT *GATA1* progenitors differentiated predominantly into CD71$^+$CD235$^+$ erythroblasts (Figure 2E, top panel). Under the same conditions, *GATA1s* progenitors produced a reduced proportion of erythroblasts, with the remainder of cells being myeloid (Figure 2E and not shown). Overall, erythroid expansion was approximately 3-fold reduced by *GATA1s* versus WT *GATA1* on a euploid background and approximately 20-fold reduced with T21 (Figure 2F). Thus, the *GATA1s* mutation overrides enhanced erythropoiesis conferred by T21. In contrast with its negative effects on erythropoiesis, *GATA1s* enhanced megakaryocyte and myeloid cell expansion in liquid cultures irrespective of chromosome 21 status (Figure 2F). These findings are consistent with previous studies showing that GATA1s mutations inhibit erythropoiesis and enhance fetal megakaryopoiesis in mouse embryos (12, 13).

Dose-dependent partial restoration of erythropoiesis by ectopic GATA1fl and GATA1s. In mice, mutations that decrease the expression of normal GATA1 protein impair erythropoiesis (25, 26). Thus, diminished levels of overall GATA1 expression may explain the

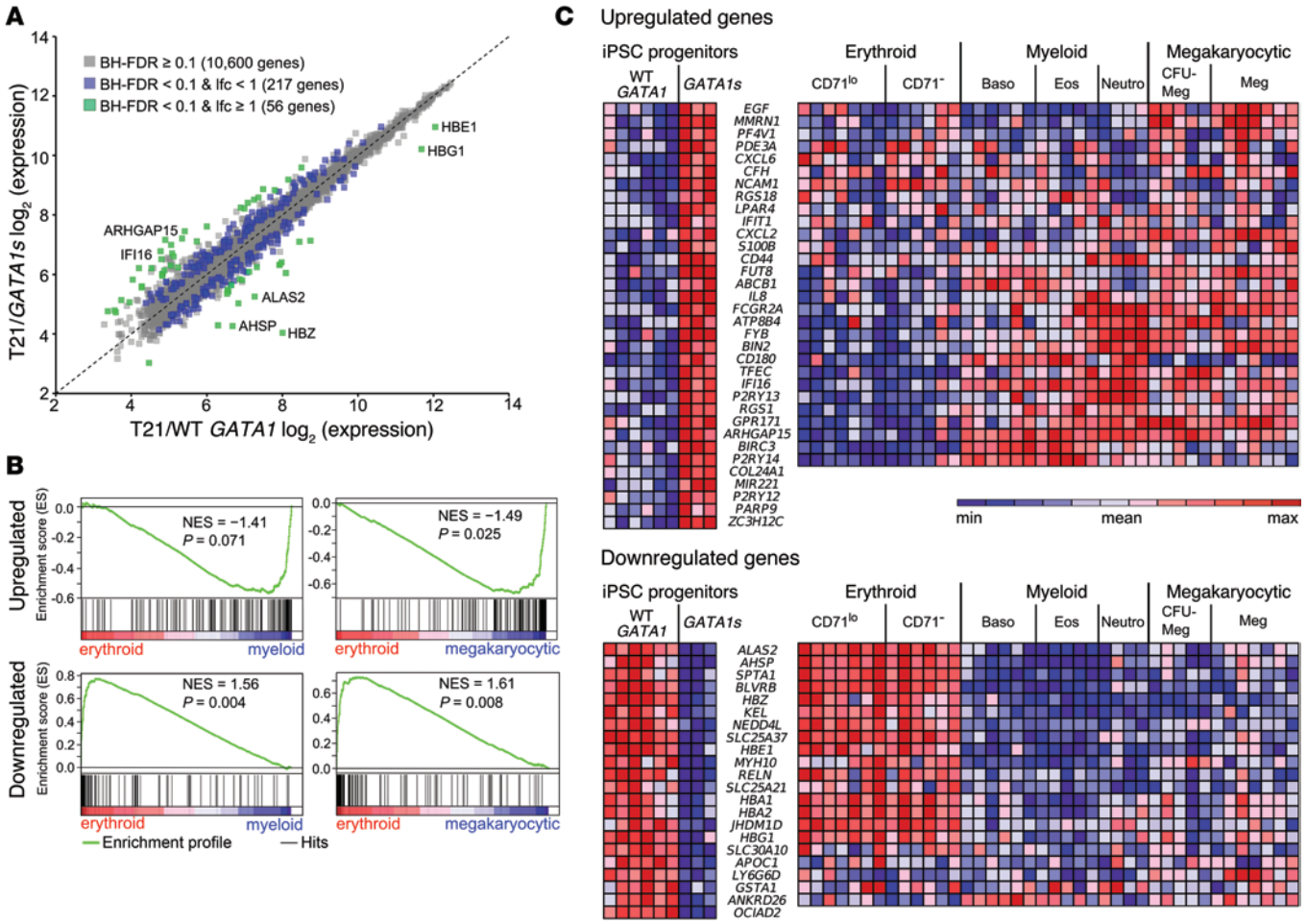

Figure 4. Global transcriptome analysis demonstrates that GATA1s mutations downregulate erythroid and upregulate myelo-megakaryocytic genes. (**A**) Mean expression values of 10,873 expressed genes in T21/*GATA1s* (n = 3 replicates) versus T21/WT *GATA1* (n = 6 replicates) iPSC–derived CD43$^+$CD41$^+$CD235$^+$ progenitors. 273 genes were differentially expressed between T21/*GATA1s* and T21/WT *GATA1*, with a fold change of mean expression of less than 2 (217 genes, blue) or 2-fold or greater (56 genes, green). (**B**) GSEA on 273 differentially expressed genes using erythroid, megakaryocytic, and myeloid expression profiles from Novershtern et al. (30). Top: enrichment of 154 genes that were upregulated in T21/*GATA1s* as compared with T21/WT *GATA1* progenitors in myeloid versus erythroid signature genes and in megakaryocytic versus erythroid genes. Bottom: enrichment of 119 downregulated genes in erythroid versus myeloid genes and in erythroid versus megakaryocytic genes. NES, normalized enrichment score; P values shown are from modified Kolmogorov-Smirnov test as implemented in GSEA. (**C**) Heat maps showing expression levels of 2-fold or greater upregulated (top) and downregulated (bottom) genes in T21/*GATA1s* versus T21/WT *GATA1* progenitors (left) as well as in lineage-committed cells (right) based on expression levels in erythroid (n = 7 replicates of CD34$^-$CD71^{lo}GlyA$^+$, 6 of CD34$^-$CD71$^-$GlyA$^+$ cells), myeloid (n = 6 replicates of basophils, 5 of eosinophils, 4 of neutrophils), and megakaryocytic (n = 5 replicates of CFU-megakaryocytes, CD34$^+$CD41$^+$CD61$^+$CD45$^-$, 7 of mature megakaryocytes, CD34$^-$CD41$^+$CD61$^+$CD45$^-$) cells from Novershtern et al. (30). Color scheme is row normalized from blue to red corresponding to minimum to maximum expression values in a given row, respectively. *ZC3H12C*, *COL24A1*, *MIR221*, *P2RY12*, *PARP9*, and *OCIAD2* were not represented on microarrays from Novershtern et al. (30).

erythropoietic defect of *GATA1s* iPSCs. To test this, we infected CD43$^+$CD41$^+$CD235$^+$ hematopoietic progenitors generated from isogenic WT *GATA1* and *GATA1s* iPSCs with lentivirus expressing GFP and GATA1fl or GATA1s cDNAs. The infected cells were cultured in erythroid-promoting cytokines, as in Figure 2, E and F. Most (~80%) of the transduced WT *GATA1* progenitors differentiated into CD235$^+$CD71$^+$ erythroblasts, independent of ectopic GATA1fl expression (Figure 3, A and B). In *GATA1s* progenitors, both GATA1fl and GATA1s overexpression partially rescued erythropoiesis, although to different extents. On average, 22% and 8% of GATA1s progenitors transduced with GATA1fl or GATA1s, respectively, differentiated into erythroblasts compared with less than 0.5% for controls. (Figure 3, A and B). In lineage-committed GATA1s mutant erythroblasts, ectopic GATA1s activated erythroid gene expression less effectively than similarly expressed GATA1fl (Figure 3, C and D). Thus, supraphysiologic levels of GATA1s (3- to 6-fold above normal, Figure 3C) can partially compensate for loss of the N-terminus to promote erythropoiesis, but are still less effective than GATA1fl.

WT GATA1 and GATA1s hematopoietic progenitor populations demonstrate transcriptional bias toward distinct lineages. To investigate how loss of the GATA1 N-terminus affects gene expression, we used the Affymetrix HuGene 1.0 ST chip to compare the transcriptomes of flow-cytometry–purified T21/*GATA1s* and T21/WT *GATA1* iPSC-derived progenitor populations (CD43$^+$CD41$^+$CD235$^+$). Among 19,392 genes examined, 10,873 were expressed in cells of one or both genotypes. A total of 273 genes were differentially

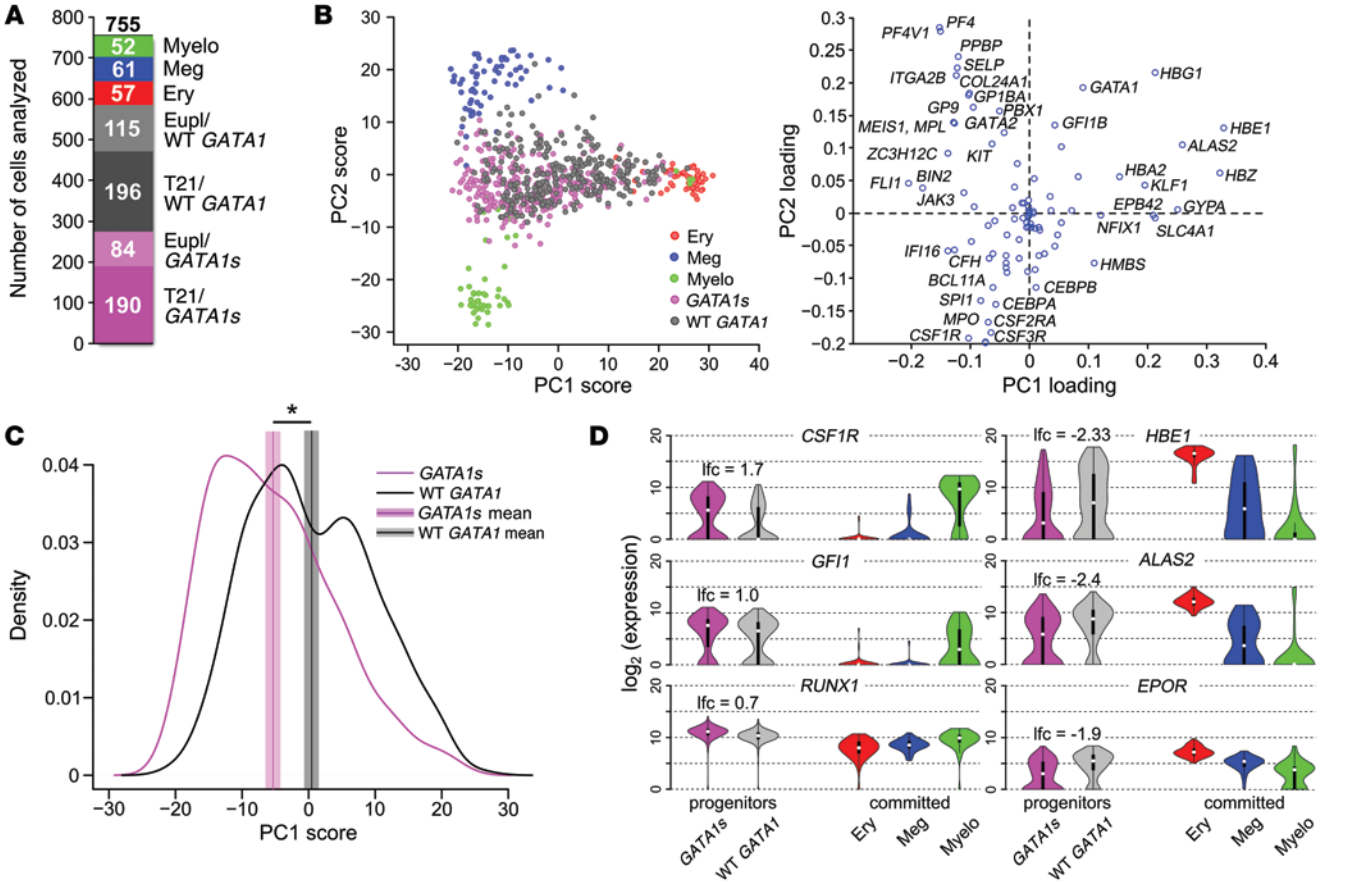

Figure 5. PCA identifies lineage signatures and reveals heterogeneity within progenitor cell populations. Single cells within CD43$^+$CD41$^+$CD235$^+$ hematopoietic progenitor populations derived from WT and mutant iPSCs were purified by flow cytometry and analyzed for gene expression by qPCR. For comparison, committed erythroid (CD41$^-$CD235$^+$), myeloid (CD45$^+$CD18$^+$), and megakaryocytic (CD41$^+$CD42$^+$) cells generated from WT iPSCs were purified and analyzed in parallel. (**A**) Numbers of analyzed single cells of each genotype (total 755 single cells). WT lineage committed cells: myeloid (myelo), megakaryocytic, and erythroid (ery). CD43$^+$CD41$^+$CD235$^+$ progenitor genotypes: euploid/WT GATA1, euploid/GATA1s, T21/WT GATA1, T21/GATA1s. (**B**) PCA on 170 committed erythroid (red), myeloid (green), and megakaryocytic (blue) cells based on expression patterns of 94 genes (right plot). PC loadings obtained from this analysis (right) were used to project 274 *GATA1s* (purple) and 311 WT *GATA1* (black) progenitor cells onto PC1 and PC2 (left). (**C**) Smooth kernel density estimate of PC1 scores of *GATA1s* (purple line) and WT *GATA1* progenitors (black line). Purple and black vertical lines represent mean PC1 scores. Shaded areas around the mean correspond to 90% confidence interval for the mean. *$P = 10^{-13}$ (Mann-Whitney *U* test). (**D**) Violin plots showing distributions of single-cell expression levels of 3 genes that were upregulated (left) and 3 genes that were downregulated (right) in *GATA1s* vs. WT *GATA1* progenitors in 5 cell types analyzed (FDR < 0.05; Mann-Whitney *U* test followed by BH-FDR correction). lfc, lfc of mean expression between *GATA1s* and WT *GATA1* progenitors.

expressed (Benjamini-Hochberg false discovery rate [BH-FDR] < 0.1; moderated *t* test followed by BH correction for multiple tests), of which 56 differed by mean expression of 2-fold or more ($\log_2$ fold change [lfc] ≥ 1, Figure 4A). Among these genes, 22 were downregulated and 34 were upregulated in *GATA1s* versus WT *GATA1* progenitors (Figure 4C and Supplemental Table 2). Transcripts that were reduced in *GATA1s* progenitors consisted mainly of erythroid genes, including those that participate in hemoglobin synthesis (*HBZ*, *HBG1*, *HBE1*, *AHSP/ERAF*, and *ALAS2*). Among these downregulated genes, 19 out of 22 were likely GATA1 targets, as indicated by ChIP-seq of human peripheral blood– and fetal liver–derived erythroblasts (Supplemental Table 2) (PBDE and PBDEFetal, The ENCODE Project; ref. 27). This frequency is about twice the value expected by chance (binomial test, $P = 10^{-4}$). Enrichment for GATA1 binding around these genes indicated that their reduced expression in *GATA1s* progenitors might be a direct consequence of GATA1 truncation. Thus, GATA1s may fail to efficiently bind or activate erythroid gene targets that are normally regulated by WT GATA1. Although 20 out of 34 genes upregulated in T21/*GATA1s* cells were bound by GATA1 in human PBDE and/or PBDEFetal cells (Supplemental Table 2), this was not significantly different from random expectation (1.3-fold enrichment; binomial test, $P = 0.084$).

We used Gene Set Enrichment Analysis (GSEA) (28, 29) to examine more broadly the lineage specificity of differentially expressed genes. Specifically, we compared the sets of 119 downregulated and 154 upregulated genes (the 273 differentially expressed genes at BH-FDR < 0.1) in *GATA1s* progenitors, with published lists of genes ordered by their differential expression in distinct hematopoietic lineages (30). The set of genes downregulated in *GATA1s* mutant progenitors was preferentially expressed in erythroid as compared with myeloid and megakaryocytic lineages ($P < 0.05$; Figure 4, B and C, and Supplemental Figure 4). Genes upregulated in *GATA1s* progenitors were significantly enriched for megakaryocytic ($P < 0.05$; Figure 4, B and C, and

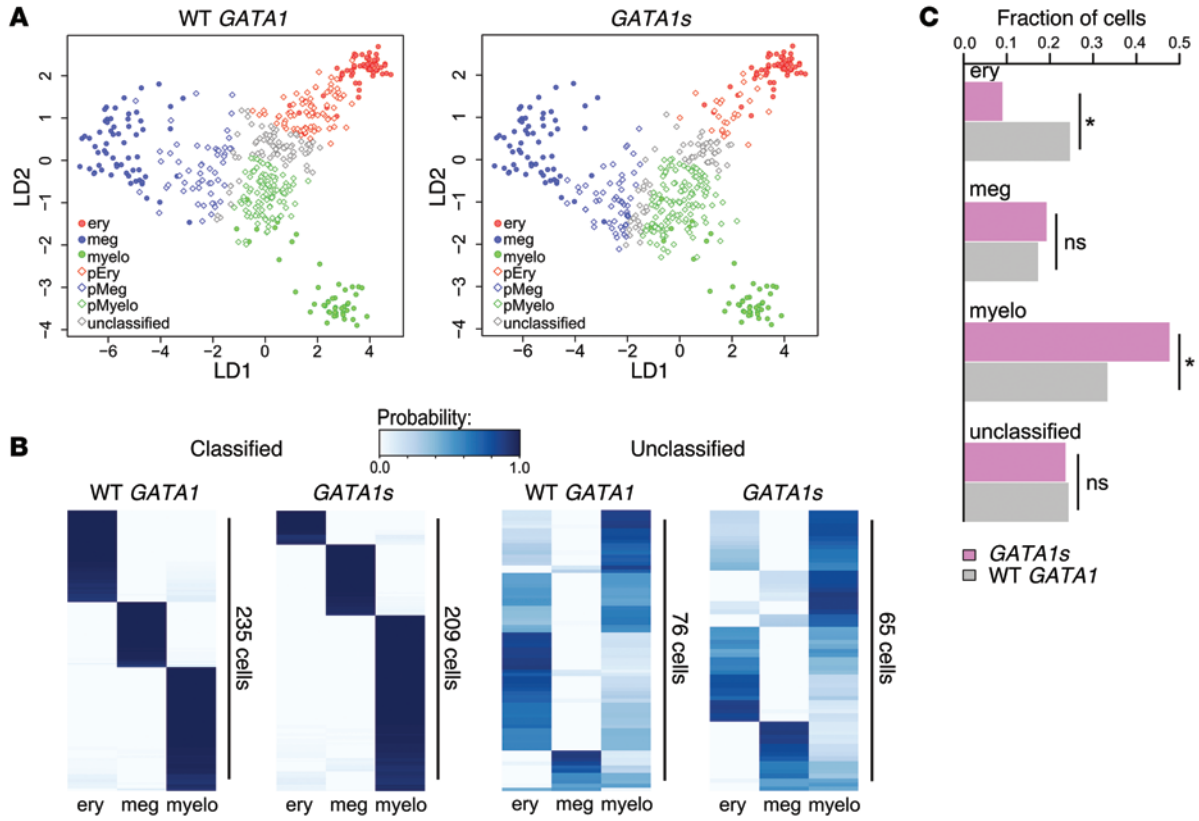

Figure 6. Single-cell gene expression analysis predicts erythroid-to-myeloid fate bias in *GATA1s* mutant progenitors. (**A**) Projections of expression patterns of iPSC-derived committed cells (filled circles in both plots) and CD43$^+$CD41$^+$CD235$^+$ progenitors expressing WT GATA1 (left, open diamonds) or GATA1s (right, open diamonds) onto linear discriminants 1 and 2 (LD1 and LD2). Projections for each cell are colored according to the classification obtained from LDA using a probability threshold of greater than 0.90. Classifications of committed cells: erythroid (red circles), megakaryocytic (blue circles), myeloid (green circles); classifications of WT *GATA1* (left) or *GATA1s* (right) progenitors: predicted erythroid (pEry) (red diamonds), predicted megakaryocytic (pMeg) (blue diamonds), predicted myeloid (pMyelo) (green diamonds). Gray diamonds represent progenitors that were unclassified at a probability threshold of greater than 0.90. (**B**) Hierarchical clustering on probabilities of belonging to erythroid, megakaryocytic, or myeloid class assigned by LDA to progenitor cells. WT *GATA1* or *GATA1s* progenitors that were assigned to a given class with a probability of greater than 0.90 are represented on 2 heat maps on the left, while cells that were unclassified (probability < 0.90) are shown on 2 heat maps on the right. (**C**) Classification of CD43$^+$CD41$^+$CD235$^+$ progenitors using LDA. Plotted are fractions of WT *GATA1* and *GATA1s* progenitor populations (consisting of a total of 311 and 274 single cells, respectively) classified into predicted cell fate with a probability of greater than 0.90. The remaining cells are plotted as unclassified. *P < 0.05 (Fisher's exact test).

Supplemental Figure 4) and myeloid lineages (P = 0.071; Figure 4, B and C; P = 0.006 for GSEA performed on 34 genes that were ≥ 2-fold upregulated, Supplemental Figure 4). GATA1 contributes to lineage bias of multipotent progenitors by activating erythroid and megakaryocytic targets while repressing a myeloid transcriptional program (reviewed in ref. 31). These observations suggest that in hematopoietic progenitor cells, GATA1s downregulates an erythroid transcriptional program and induces a myelo-megakaryocytic program. Transcriptome comparison of euploid *GATA1s* versus WT *GATA1* iPSC–derived progenitors demonstrated a similar pattern, suggesting that these effects of GATA1s are independent of T21 (Supplemental Figure 5).

Single-cell gene expression analysis reveals heterogeneity in WT GATA1 and GATA1s progenitor populations. Our analysis of transcriptomes in progenitor populations showed that *GATA1s* mutations confer a strong bias toward myelo-megakaryocytic transcription. However, individual phenotypically matched stem and progenitor cells exhibit considerable heterogeneity in gene expression (32, 33). Thus, subpopulations of cells within our iPSC-derived progenitor pools could exhibit different fate biases. To investigate this, we used quantitative real-time PCR (qPCR) to compare expression levels of 91 selected hematopoietic genes and 3 housekeeping genes (Supplemental Table 3) in 755 individual, flow-cytometry–purified CD43$^+$CD41$^+$CD235$^+$ iPSC–derived progenitors. Input cell genotypes included *GATA1s* (84 euploid + 190 T21 = 274 cells) and WT *GATA1* (115 euploid + 196 T21 = 311 cells) iPSC–derived progenitors. As controls, we examined euploid/WT *GATA1* lineage–committed erythroid (CD41$^-$CD235$^+$, n = 57), megakaryocytic (CD41$^+$CD42$^+$, n = 61), and myeloid (CD45$^+$CD18$^+$, n = 52) cells (Figure 5A).

The expression profiles of the single cells can be represented as 755 data points (one per cell) in a 94-dimensional gene space. We used principal component analysis (PCA) to identify weighted combinations of gene expression that capture the major differences among erythroid, megakaryocytic, and myeloid lineages. Each weighted combination is a principal component (PC), which is a vector composed of 94 coefficients (also called loadings) related to the contribution of each gene's expression to the PC (Figure 5B).

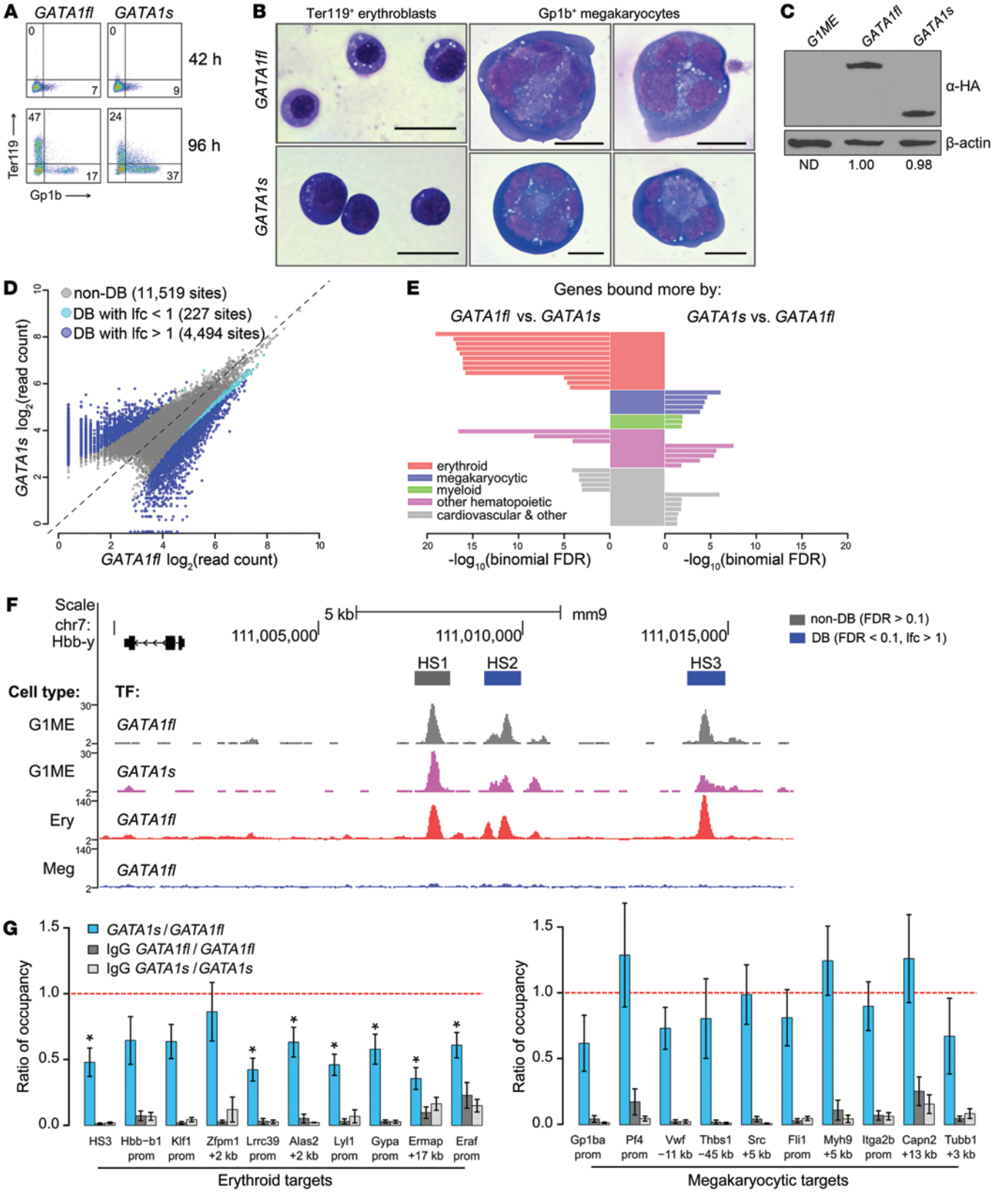
A
GATA1fl
GATA1s
42 h
96 h
Ter119
Gp1b
B
Ter119+ erythroblasts
Gp1b+ megakaryocytes
GATA1fl
GATA1s
C
G1ME
GATA1fl
GATA1s
α-HA
β-actin
ND
1.00
0.98
D
non-DB (11,519 sites)
DB with lfc < 1 (227 sites)
DB with lfc > 1 (4,494 sites)
GATA1s log2(read count)
GATA1fl log2(read count)
E
Genes bound more by:
GATA1fl vs. GATA1s
GATA1s vs. GATA1fl
erythroid
megakaryocytic
myeloid
other hematopoietic
cardiovascular & other
-log10(binomial FDR)
F
Scale
chr7:
Hbb-y
5 kb
mm9
111,005,000
111,010,000
111,015,000
non-DB (FDR > 0.1)
DB (FDR < 0.1, lfc > 1)
HS1
HS2
HS3
Cell type:
TF:
G1ME
GATA1fl
G1ME
GATA1s
Ery
GATA1fl
Meg
GATA1fl
G
GATA1s/GATA1fl
IgG GATA1fl/GATA1fl
IgG GATA1s/GATA1s
Ratio of occupancy
HS3
Hbb−b1 prom
Klf1 prom
Zfpm1 +2 kb
Lrrc39 prom
Alas2 +2 kb
Lyl1 prom
Gypa prom
Ermap +17 kb
Eraf prom
Erythroid targets
Gp1ba prom
Pf4 prom
Vwf −11 kb
Thbs1 −45 kb
Src +5 kb
Fli1 prom
Myh9 +5 kb
Itga2b prom
Capn2 +13 kb
Tubb1 +3 kb
Megakaryocytic targets

Figure 7. Loss of the N-terminus selectively inhibits GATA1 binding to erythroid genes in *Gata1*⁻ MEPs. (**A**) Flow cytometry of G1ME cells at 42 and 96 hours after transduction with HA-GATA1fl or HA-GATA1s. Numbers denote percentage of total cells in the indicated quadrant. (**B**) May-Grünwald-Giemsa staining of G1ME-derived erythroblasts (left) and megakaryocytes (right) 96 hours after transduction. Scale bars: 20 μm. (**C**) Western blot at 42 hours after transduction showing expression of HA-GATA1fl and HA-GATA1s relative to β-actin. ND, not determined. (**D**) Genome-wide ChIP-seq binding signals of GATA1fl and GATA1s at 42 hours after transduction. Plotted are mean read counts (n = 2 replicates each). non-DB, nondifferentially bound (gray dots); DB, differentially bound sites (blue and cyan; FDR < 0.1). (**E**) Functional enrichment analysis using GREAT. Plotted are significance values for top 10 mouse phenotype and GO biological process enrichment terms, classified as erythroid, megakaryocytic, myeloid, other hematopoietic, or cardiovascular and other. List of terms can be found in Supplemental Table 5. (**F**) GATA1 binding at the β-globin locus. ChIP-seq tracks, top to bottom: GATA1fl and GATA1s in G1MEs, GATA1fl in primary mouse erythroblasts and megakaryocytes (42). Rectangles above tracks: size of binding sites analyzed in differential binding analysis; gray, non-DB; blue, DB. (**G**) Anti-HA ChIP at selected GATA1-binding sites in G1MEs shown as ratio of GATA1s to GATA1fl occupancy ± SEM (n = 4). *Ratio significantly different than 1 at $P < 0.05$ (2-tailed Student's t test). IgG GATA1s and IgG GATA1fl controls for nonspecific binding.

Notably, the first 2 PCs — PC1 and PC2 — accounted for 65% of the variance in the data. Projection of the expression patterns from committed cells onto the plane of PC1 and PC2 resulted in clear clustering of the 3 reference cell types (Figure 5B), showing that PC1 and PC2 can distinguish specific lineage signatures of erythroid, megakaryocytic, and myeloid cells.

Expression profiles from individual WT *GATA1* and *GATA1s* progenitors projected onto PC1 and PC2 were distributed heterogeneously among the 3 differentiated cell lineages (Figure 5B), with some progenitor cells located closer to one of the mature cell types and others located more equally among the 3 reference cell types. This heterogeneity suggested that the PC1 and PC2 scores could also be used to estimate the likelihood of a single progenitor differentiating toward erythroid, megakaryocytic, or myeloid lineages. Although the distributions of WT *GATA1* and *GATA1s* progenitors overlapped, cells expressing GATA1s were shifted along PC1 toward myeloid and megakaryocytic reference cells relative to WT GATA1, as illustrated by density plots of PC1 scores (i.e., projections of expression patterns of each cell along PC1) (Figure 5, B and C, $P = 10^{-13}$, Mann-Whitney U test). Examples of genes that are differentially expressed in WT *GATA* versus *GATA1s* mutant progenitors are shown in Figure 5D.

We used PC1 and PC2 as lineage signatures within a linear discriminant analysis (LDA) to assign to each progenitor cell a probability of belonging to each of the 3 cell types. Biologically, these probabilities can be used to approximate the likelihood that a particular progenitor cell will differentiate toward a given lineage and thus infer its developmental potential within 4 categories: erythroid, megakaryocytic, myeloid, or unclassified (probability threshold > 0.90) (Figure 6, A and B). Compared with WT *GATA1* progenitors, cells expressing GATA1s were much less frequently classified as likely to differentiate toward erythroid ($P = 6 \times 10^{-7}$, Fisher's exact test) and more likely to be classified myeloid ($P = 5 \times 10^{-4}$; Figure 6C). In contrast, the cell-fate predictions between WT *GATA1* and *GATA1s* progenitors for megakaryocytic fate and unclassified were not significantly different ($P > 0.05$, Figure 6C). Unclassified progenitors were heterogeneous, with varying probabilistic preferences, and may represent a less-differentiated subpopulation compared with classified cells (Figure 6B). In summary, single-cell gene expression analysis suggests that *GATA1s* mutant progenitors are biased toward myeloid fates and away from an erythroid fate, relative to a surface phenotype-matched WT *GATA1* population.

Cell-fate bias predicted by gene expression studies could be executed by a few driver genes or by small contributions from many genes. We searched for potential driver genes by examining the distributions of expression levels for all 94 genes interrogated among single cells. These distributions differed dramatically among the committed myeloid, erythroid, and megakaryocytic lineages, as expected (Figure 5D and Supplemental Figure 6). Comparison of WT *GATA1* and *GATA1s* mutant progenitors identified 40 of 94 differentially expressed genes (BH-FDR < 0.05; Mann-Whitney U test followed by BH correction; Supplemental Table 4, Figure 5D, and Supplemental Figure 6). Importantly, for most of these genes (36 of 40), expression in WT *GATA1* progenitors is shifted toward that of erythroid cells, while expression in *GATA1s* progenitors favored myeloid and/or megakaryocytic fates (Figure 5D and Supplemental Figure 6). Moreover, many of the differentially expressed genes exhibited only small changes in mean expression (25 of 40 less than 2-fold change; Supplemental Table 4). Thus, the aggregate of small changes in expression of many hematopoietic genes may cause erythroid–to–myelo-megakaryocytic lineage bias of *GATA1s* progenitors.

Defective erythropoietic activity of GATA1s in murine megakaryocyte-erythroid progenitor–like cells. We compared the properties of GATA1fl and GATA1s in GATA1⁻ megakaryocyte-erythroid (G1ME) cells, a megakaryocyte-erythroid progenitor–like (MEP-like) line derived from *Gata1*⁻ ESCs (34). Four days after transduction by GATA1fl-expressing retrovirus, G1ME cells differentiated into Ter119⁺ erythroblasts and GpIb⁺ (CD42) megakaryocytes (Figure 7, A and B), as reported previously (34). However, erythropoiesis was impaired after expression of GATA1s, as evidenced by reduced proportion of Ter119⁺ cells (Figure 7A) and stalled erythroid maturation revealed by histological staining (Figure 7B). In contrast, GATA1s induced an increased proportion of normal-appearing megakaryocytes (Figure 7, A and B). Importantly, GATA1fl and GATA1s were expressed at similar levels in G1ME cells (Figure 7C; see complete unedited blots in the supplemental material). Thus, the observed differences in the hematopoietic activities of these 2 isoforms are not due to dosage discrepancies among the virally encoded proteins. Overall, these complementation studies in G1ME cells are consistent with the phenotypes of patients harboring germline *GATA1s* mutations (2, 3) and our in vitro differentiation studies of *GATA1s* patient-derived iPSCs (Figures 1 and 2).

Loss of the N-terminus selectively decreases GATA1 binding at erythroid genes. Given that other germline *GATA1* mutations outside the DNA-binding domain affect the ability of GATA1 to bind target sites in vivo (35–38), we investigated whether loss of the GATA1 N-terminus also impairs chromatin occupancy. We performed ChIP-seq on G1ME cells 42 hours after retroviral transfer of HA-tagged GATA1fl or GATA1s cDNAs. At this time point, there was no apparent difference in the cell-surface phenotypes between GATA1fl- and GATA1s-expressing cells (Figure 7A, top row).

Genome-wide analysis of 16,231 binding sites occupied by GATA1fl and/or GATA1s in G1ME cells identified 4,721 sites that were differentially bound (FDR < 0.1; edgeR analysis implemented by DiffBind package in R; refs. 39, 40) (Figure 7D). This included 2,106 sites that were bound more by GATA1fl than GATA1s and 2,615 sites that were bound more by GATA1s than GATA1fl. We used Genomic Regions Enrichment of Annotations Tool (GREAT) to investigate the functional implications of differential DNA binding by GATA1s (> 2-fold difference in binding signal) (41). Importantly, genes assigned to sites bound more by GATA1fl than GATA1s were highly enriched for terms associated with erythroid functions (Figure 7E and Supplemental Table 5). Consistent with these findings, genome-wide microarray transcriptome analysis comparing mean gene expression levels between cells expressing GATA1s versus GATA1fl at 42 hours after transduction revealed downregulation of an erythroid program and upregulation of a megakaryocytic program in GATA1s-expressing cells (Supplemental Figure 7). We confirmed the decreased binding of GATA1s relative to GATA1fl at key erythroid regulatory regions by ChIP-qPCR in independent transductions of G1ME cells (Figure 7G) and observed decreased expression of erythroid GATA1 target genes (Supplemental Figure 8). Thus, we conclude that the erythropoietic defect observed for the loss of the GATA1 NAD results at least in part from decreased binding of GATA1s at selected sites regulating erythroid genes.

In contrast, the increased binding associated with loss of the GATA1 NAD occurs almost exclusively at sites with weak binding by GATA1s (Figure 7D and Supplemental Figure 9). While the presumptive target genes assigned to those sites were enriched for nonerythroid terms, including megakaryocytic and myeloid, these enrichments had lower statistical significance than erythroid targets associated with decreased binding of GATA1s (Figure 7E). Restricting the analysis to sites with stronger binding reduces the number of terms enriched for the presumptive targets (Supplemental Figure 9). Furthermore, the ChIP-qPCR results showed nonsignificant differences between GATA1fl and GATA1s binding at a subset of known megakaryocytic regulatory regions (Figure 7G), although expression of some of the tested megakaryocytic genes increased (Supplemental Figure 8).

While GATA1s binds differentially relative to GATA1fl at many sites, we observed equivalent occupancy by GATA1fl and GATA1s at many other sites, including some sites regulating erythroid genes. This demonstrates that the N-terminus facilitates DNA interactions selectively, even within the same lineage. For example, GATA1s binding was decreased at the HS2 and HS3 regulatory regions of the Hbb-y locus compared with GATA1fl (FDR < 0.1), but was equivalent at the HS1 site (Figure 7F). These β-globin locus target sites are not bound by GATA1fl in megakaryocytes (ChIP-seq track from ref. 42). Together, these data show that loss of the GATA1 N-terminus decreases chromatin occupancy at specific sites within erythroid target genes, which likely contributes to erythropoietic deficiencies mediated by the GATA1s protein.

Discussion

GATA1s mutations that remove amino acids 1–83 of the full-length GATA1 protein via alternate splicing and use of an internal initiation codon were first identified in DS-associated AMKL and TMD (5). More recently, similar mutations were discovered to cause anemia in gene-targeted mouse embryos (12, 13) and in euploid patients with hypoplastic anemia, including some who were assigned a diagnosis of DBA (2, 3). Here, we used patient-derived iPSCs and mouse ESCs to confirm that *GATA1s* mutations impair erythropoiesis and investigated the underlying mechanisms. A recent study used gene-editing and chromosome-transfer techniques to show that a *GATA1s* mutation inhibits the erythropoietic capacity of euploid and T21 human ESCs (43). Our current findings confirm this result in iPSCs generated from multiple patients and identify underlying mechanisms. Specifically, we show that loss of the amino-terminus decreases binding to selective sites within erythroid genes and biases hematopoietic output toward cells of megakaryocytic and myeloid fates at the expense of erythropoiesis.

GATA1 regulates hematopoietic cell fates by activating lineage-specific genes and repressing those of alternate cell types (31). Previous loss-of-function and overexpression studies demonstrated that GATA1 stimulates differentiation of multipotent progenitors toward erythroid fate (44–48) and also acts as a survival/maturation factor for committed erythroblasts (26, 49). The reduced erythropoietic activities of GATA1s demonstrated in the current study could occur through either or both mechanisms. Experiments in bipotential G1ME cells indicate that GATA1s is deficient in both promotion of erythroid differentiation of bipotential progenitors (Figure 7A) and maturation of committed erythroid progenitors (Figure 7B) compared with GATA1fl. It is more difficult to distinguish between these mechanisms in iPSC differentiation cultures, since $CD43^+CD41^+CD235^+$ progenitors generally give rise to relatively few multilineage colonies (refs. 15, 50, 51, and Figure 2A), which limited our ability to investigate the effects of GATA1s on lineage determination. Thus, we cannot ascribe the myelo-megakaryocytic lineage bias observed for the GATA1s $CD43^+CD41^+CD235^+$ population unambiguously to either a shift in lineage choice or altered survival/proliferation of monolineage-committed progenitors. We did find that WT *GATA1* and *GATA1s* iPSCs produce similar numbers of $CD43^+CD41^+CD235^+$ progenitors, and thus, if the loss of erythroid potential resulted from a decrease in erythroid-committed progenitors, then the myelo-megakaryocytic progenitors replaced them. The lineage signatures derived from expression analysis of individual progenitor cells reveal a subpopulation of cells that are not readily classified as likely to differentiate to 1 of the 3 lineages. These less-differentiated progenitors could have multilineage potential, but such an assignment awaits formal fate-mapping studies.

Both WT GATA1 and GATA1s produce dosage-dependent effects (9, 25, 26). *GATA1s* mutations typically reduce overall GATA1 protein output. Thus, it is possible that GATA1s functions normally, but is expressed at insufficient levels for optimal erythropoiesis. Supporting this, mutant mice expressing modestly reduced levels of WT GATA1 protein are anemic (25, 26), and transgenic rescue with high levels of GATA1s restores erythropoiesis in a different strain of *Gata1*-deficient mice (9). Moreover, we show that enforced expression of ectopic GATA1s enhances erythropoiesis in iPSC-derived progenitors from patients with *GATA1s* mutations (Figure 3). Alternatively, the GATA1 N-terminus may mediate unique functions specifically in erythroid cells. In support, ectopic GATA1s is inferior to an equivalent level of GATA1fl for promoting erythropoiesis in *GATA1s* iPSC–derived hematopoietic progenitors (Figure 3) and in bipotential *Gata1*$^-$ G1ME cells (Figure 7).

Considering our current findings and prior studies of others, *GATA1s* mutations likely cause anemia both by reducing overall GATA protein expression and by favoring the production of a functionally impaired form of the transcription factor. Our work provides mechanistic insights into the latter by showing that the loss of GATA1 N-terminus decreases binding to erythroid target genes. *GATA1s* mutations could inhibit erythropoiesis by altering the expression of 1 key lineage-determining gene or through aggregate effects on multiple genes. Supporting the latter, we show that individual mutant progenitor cells exhibit small but significant changes (BH-FDR < 0.05) in the expression of multiple key lineage regulators, including *KLF1*, *GATA2*, *RUNX1*, and *CEBPA*.

How the N-terminus regulates GATA1 chromatin occupancy to enhance erythroid gene expression represents an interesting problem for future studies. GATA1 binds DNA directly via subdomains of its 2 zinc fingers, most importantly the C-finger (52, 53). However, several disease-associated amino acid substitutions within the N-finger inhibit GATA1 binding to target genes in vivo, yet have no effect on in vitro binding to naked DNA templates. These regions of the N-finger do not interact with DNA directly, but rather are believed to promote GATA1 chromatin occupancy by recruiting essential cofactors, including FOG1 and the TAL1/LMO2/LDB1 complex (35, 38). Similarly, loss of the Kruppel-like factor 3 transcription N-terminus, which also does not bind DNA directly, alters chromatin occupancy selectively at a subset of in vivo target genes in mouse embryonic fibroblasts (54). Taken together, our current findings suggest that interactions between currently unidentified cofactors and the GATA1 N-terminus facilitate chromatin occupancy specifically at erythroid genes. RUNX1 and the retinoblastoma protein (Rb) bind the GATA1 N-terminus (55, 56). However, loss of RUNX1 in hematopoietic progenitors impairs megakaryopoiesis, but not erythropoiesis (reviewed in ref. 57). Rb facilitates terminal maturation of erythroid precursors (reviewed in ref. 58), but has no known role in their lineage commitment. Thus, additional proteins likely modulate the erythropoietic functions of GATA1 through interactions with its N-terminus.

GATA1s mutations have overlapping and distinct effects on hematopoiesis in humans and mice. In mice, *Gata1s* mutations inhibit primitive-type (yolk sac derived) erythropoiesis, resulting in partial embryonic lethality (12, 13), but postnatal hematopoiesis is normal. In contrast, similar mutations in humans cause erythroid hypoplasia and anemia throughout life (2, 3). In the current study, we used iPSC in vitro hematopoietic differentiation protocols that generate erythroid cells expressing mainly embryonic type globins (ζ and ε), thereby resembling yolk sac–type hematopoiesis. These data, taken together with murine studies, suggest that some human embryos with *GATA1s* mutations may not survive to birth. Indeed, classical DBA caused by heterozygous mutations in ribosomal protein genes can cause hydrops fetalis and in utero demise (59, 60). These clinical observations predict that iPSCs from *GATA1s* mutant patients will exhibit defective fetal liver/bone marrow–type "definitive" erythropoiesis. Currently, methods to produce adult-type definitive erythroblasts from human iPSCs are suboptimal, particularly for quantitative comparative studies between different mutant and control lines. However, our preliminary experiments indicate that EBs derived from WT human iPSCs, but not *GATA1s* iPSCs, generate erythroblasts expressing definitive-type globins (γ and β) after prolonged culture (> 30 days) (not shown). The enhanced myelopoiesis and megakaryopoiesis observed in *GATA1s* iPSCs regardless of karyotype diverge from phenotypes of individuals with germline *GATA1s* mutations that are associated with anemia and neutropenia. Induction of adult-type "definitive" hematopoiesis in iPSC cultures may be required to better recapitulate all clinical phenotypes. Overall, our current findings show that patient-derived iPSCs recapitulate many clinical aspects of germline *GATA1s* mutations and provide a model system for better defining the role of the GATA1 N-terminus in human erythropoiesis.

While the current study focuses on the role of *GATA1s* mutations in erythropoiesis, our findings also provide insights into how these mutations interact with T21. Previous work has shown that T21 augments erythropoiesis in primary fetal cells (61–63) and iPSCs (18, 64). Here, we confirm these findings in T21/WT *GATA1* iPSCs and demonstrate that GATA1s overrides the erythropoietic effects of T21 in isogenic double-mutant T21/*GATA1s* iPSCs derived from 2 different patients with TMD. In addition, we show that GATA1s enhances human megakaryocyte and myeloid expansion, which likely contributes to the development of DS-associated myeloproliferative disorders. A LXCXE motif within the N-terminus of GATA1 (amino acids 81–85) binds Rb-E2F transcription factor family members, which inhibits the expression of genes that stimulate cell proliferation (56, 65). Studies of *GATA1* alleles in DS-TMD patients, including rare cases with mutations that produce internal protein deletions (66), map a minimal disease-associated domain to within amino acids 74–83, which includes the Rb-binding motif. It will be interesting to determine whether this region also facilitates erythroid differentiation or, alternatively, whether this effect can be uncoupled from the antiproliferative actions of GATA1 on myelopoiesis and megakaryopoiesis via distinct mutations affecting the N-terminus.

Methods

Generation of iPSCs. Mononuclear cells were infected with pHage2-CMV-RTTA-W and pHage-Tet-hSTEMMCA-loxP virus with human *OCT4*, *SOX2*, *KLF4*, or *MYC*, as previously performed (17, 18). Reprogramming details and characterization methods are provided in Supplemental Methods.

Hematopoietic differentiation. iPSC EB formation, hematopoietic colony assays, liquid culture assays, lentiviral infection, and flow cytometry were performed as previously described (18) as and described in Supplemental Methods.

Cell line culture. G1ME cells were generated, cultured, and genetically manipulated as described by our laboratory (34). The MSCV-based retroviral vector MIGR1-GFP was used to express HA-tagged murine GATA1 or GATA1s in G1ME cells.

Western blot. Nuclear extracts were prepared according to standard methods, fractionated on sodium dodecyl sulfate–2% polyacrylamide gels, and transferred to nitrocellulose membranes by electroblotting. Antibodies included anti-human GATA1 (ab11852, Abcam), HA (sc-805, Santa Cruz Biotechnology Inc.), and actin (A3854, Sigma-Aldrich). Band intensities were quantified using ImageJ (http://imagej.nih.gov/ij/).

qPCR. RNA was isolated with Trizol (Invitrogen) or the RNeasy kit (QIAGEN), complementary DNA prepared by the oligo(dT) method (Invitrogen), and PCR quantified using SYBR green (Roche) on a ViiA7 (Life Technologies). Expression was normalized to cyclophilin, *Gapdh*, or β-actin, and relative quantification determined by the comparative Ct method. See Supplemental Methods for primer sequences.

Microarray transcriptome analysis. RNA was isolated and hybridized to Affymetrix HuGene 1.0 ST (human iPSC-derived progenitors) or Affymetrix Mouse Genome 430 2.0 Microarrays (mouse G1ME cells at 42 hours after transduction with GATA1 or GATA1s). To obtain gene expression values, raw intensity values from Affymetrix CEL data files were processed using the robust multichip average (RMA) method (67), implemented by the Bioconductor R "oligo" (HuGene 1.0 ST arrays) or "affy" (Mouse Genome 430 2.0 arrays) package (68–70). All original microarray data were deposited in the NCBI's Gene Expression Omnibus (GEO GSE61933). See Supplemental Methods for details on bioinformatics analysis.

Single-cell gene expression analysis. Preamplified cDNA from sorted single cells was loaded onto the Fluidigm BioMark 96.96 Dynamic Array chips according to the manufacturer's instructions. See Supplemental Methods for details on data processing and bioinformatics analysis as well as Supplemental Table 6 for raw and normalized single-cell gene expression data.

ChIP. ChIP was performed as described (71) with anti-HA (sc-805; Santa Cruz Biotechnology Inc.) antibody. For ChIP-qPCR, DNA was quantified by qPCR with SYBR Green on a ViiA7. See Supplemental Methods for primer sequences. ChIP-seq libraries were prepared as outlined for Illumina's TruSeq ChIP Sample Prep Kit (IP-202-1012), except that the libraries were size selected using Agencourt SPRIselect Beads for an average size of approximately 300 to 325 bp prior to PCR amplification. Libraries were sequenced on the Illumina HiSeq2000. Sequence reads were mapped to the reference mouse genome (NCBI37/mm9, July 2007) using Bowtie 1.0.0. ChIP-seq peaks were called using MACS 1.3.7.1. All ChIP-seq data were deposited in the NCBI's Gene Expression Omnibus (GEO GSE61933). See Supplemental Methods for details on differential binding analysis.

Statistics. Statistical analyses were performed using a 2-tailed Student's *t* test (Figures 1 and 2 and Figure 7G). Genome-wide microarray differential expression analyses were performed using a moderated *t* test (72). Statistical analysis for GSEA was performed as described (29). Statistical analyses of comparisons between non-normal distributions of data were performed using nonparametric Mann-Whitney *U* test (Figure 5, C and D). The significance of differences in counts was assessed by Fisher's exact test (Figure 6C). Differential binding analysis was assessed using edgeR (39, 40). Results were considered statistically significant if the *P* value was less than 0.05. BH correction for multiple testing (73) was applied to microarray differential expression analyses, single-cell gene expression analysis, and differential binding analysis.

Study approval. All patients or their families provided written informed consent to participate in this study. The institutional review board at the Children's Hospital of Philadelphia approved the study protocol.

Provision of materials. The vector pHage-Tet-hSTEMMCA-loxP was a gift from Darrell Kotton (Boston University, Boston, Massachusetts, USA) and Gustavo Mostoslavsky (Boston University, Boston, Massachusetts, USA).

Acknowledgments

We thank the Children's Hospital of Philadelphia Human ESC/iPSC Core Facility located in the Center for Cellular and Molecular Therapeutics for technical and scientific advice, and the Flow Cytometry and Molecular Profiling core facilities at the University of Pennsylvania as well as Cheryl A. Keller and Belinda M. Giardine from the Center for Comparative Genomics and Bioinformatics at The Pennsylvania State University for their help with ChIP-seq experiments and ChIP-seq data processing. This work was supported by NIH K08 HL093290 and NIH R01 DK100854 (to S.T. Chou), NIH RC2 HL10166 (to M.J. Weiss), NIH P30 DK090969 (to M.J. Weiss and S.T. Chou), and NIH R01 DK065806 (to R.C. Hardison), and an American Society of Hematology Scholar Award and Alex's Lemonade Stand Foundation Springboard Grant (to S.T. Chou).

Address correspondence to: Stella T. Chou, The Children's Hospital of Philadelphia, 3615 Civic Center Blvd., 316 ARC, Philadelphia, Pennsylvania 19104, USA. Phone: 215.590.0947; E-mail: chous@email.chop.edu.

1. 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 U S A*. 1995;92(25):11598–11602.
2. Hollanda LM, 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.
3. Sankaran VG, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest*. 2012;122(7):2439–2443.
4. Parrella S, et al. Loss of GATA-1 full length as a cause of Diamond-Blackfan anemia phenotype. *Pediatr Blood Cancer*. 2014;61(7):1319–1321.
5. Wechsler J, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet*. 2002;32(1):148–152.
6. Alford KA, et al. Analysis of GATA1 mutations in Down syndrome transient myeloproliferative disorder and myeloid leukemia. *Blood*. 2011;118(8):2222–2238.
7. 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.
8. 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.
9. 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.
10. Kuhl C, Atzberger A, Iborra F, Nieswandt B, Porcher C, Vyas P. GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. *Mol Cell Biol*. 2005;25(19):8592–8606.
11. 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(22):12355–12358.
12. Li Z, Godinho FJ, Klusmann JH, 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.
13. Birger Y, et al. Perturbation of fetal hematopoiesis in a mouse model of Down syndrome's transient myeloproliferative disorder. *Blood*. 2013;122(6):988–998.
14. Woo AJ, et al. Developmental differences in IFN signaling affect GATA1s-induced megakaryocyte hyperproliferation. *J Clin Invest*. 2013;123(8):3292–3304.
15. Kennedy M, D'Souza SL, Lynch-Kattman M, Schwantz S, Keller G. Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood*. 2007;109(7):2679–2687.
16. Niwa A, et al. A novel serum-free monolayer culture for orderly hematopoietic differentiation of human pluripotent cells via mesodermal progenitors. *PLoS One*. 2011;6(7):e22261.
17. Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells*. 2009;27(3):543–549.
18. Chou ST, et al. Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc Natl Acad Sci U S A*. 2012;109(43):17573–17578.
19. Anguera MC, et al. Molecular signatures of

human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell Stem Cell*. 2012;11(1):75–90.
20. Tchieu J, et al. Female human iPSCs retain an inactive X chromosome. *Cell Stem Cell*. 2010;7(3):329–342.
21. Ananiev G, Williams EC, Li H, Chang Q. Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from Rett syndrome patients as in vitro disease model. *PLoS One*. 2011;6(9):e25255.
22. Mekhoubad S, Bock C, de Boer AS, Kiskinis E, Meissner A, Eggan K. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell*. 2012;10(5):595–609.
23. Liang G, Zhang Y. Genetic and epigenetic variations in iPSCs: potential causes and implications for application. *Cell Stem Cell*. 2013;13(2):149–159.
24. Mills JA, et al. Clonal genetic and hematopoietic heterogeneity among human-induced pluripotent stem cell lines. *Blood*. 2013;122(12):2047–2051.
25. McDevitt MA, Shivdasani RA, Fujiwara Y, Yang H, Orkin SH. A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proc Natl Acad Sci U S A*. 1997;94(13):6781–6785.
26. Pan X, et al. Graded levels of GATA-1 expression modulate survival, proliferation, and differentiation of erythroid progenitors. *J Biol Chem*. 2005;280(23):22385–22394.
27. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
28. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267–273.
29. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
30. Novershtern N, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296–309.
31. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*. 2008;132(4):631–644.
32. Narsinh KH, et al. Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. *J Clin Invest*. 2011;121(3):1217–1221.
33. Pina C, et al. Inferring rules of lineage commitment in haematopoiesis. *Nat Cell Biol*. 2012;14(3):287–294.
34. 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.
35. Campbell AE, Wilkinson-White L, Mackay JP, Matthews JM, Blobel GA. Analysis of disease-causing GATA1 mutations in murine gene complementation systems. *Blood*. 2013;121(26):5218–5227.
36. 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.
37. Letting DL, Chen YY, Rakowski C, Reedy S, Blobel GA. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci U S A*. 2004;101(2):476–481.
38. Pal S, et al. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A*. 2004;101(4):980–985.
39. Stark R, Brown G. Diffbind: differential binding analysis of ChIP-seq peak data. Bioconductor Web site. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf. Published November 27, 2014. Accessed January 7, 2015.
40. Ross-Innes CS, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481(7381):389–393.
41. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495–501.
42. Pimkin M, et al. Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res*. 2014;24(12): 1932–1944.
43. Kazuki Y, et al. Down syndrome-associated haematopoiesis abnormalities created by chromosome transfer and genome editing technologies. *Sci Rep*. 2014;4:6136.
44. May G, et al. Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. *Cell Stem Cell*. 2013;13(6):754–768.
45. Heyworth C, Pearson S, May G, Enver T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *EMBO J*. 2002;21(14):3770–3781.
46. Rhodes J, et al. Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev Cell*. 2005;8(1):97–108.
47. Mancini E, et al. FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. *EMBO J*. 2012;31(2):351–365.
48. Arinobu Y, et al. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myloerythroid and myelolymphoid lineages. *Cell Stem Cell*. 2007;1(4):416–427.
49. Weiss MJ, Orkin SH. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc Natl Acad Sci U S A*. 1995;92(21):9623–9627.
50. Klimchenko O, et al. A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis. *Blood*. 2009;114(8):1506–1517.
51. Vodyanik MA, Bork JA, Thomson JA, Slukvin II. Human embryonic stem cell-derived CD34$^+$ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood*. 2005;105(2):617–626.
52. Ko LJ, Engel JD. DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol*. 1993;13(7):4011–4022.
53. Merika M, Orkin SH. DNA-binding specificity of GATA family transcription factors. *Mol Cell Biol*. 1993;13(7):3999–4010.
54. Burdach J, et al. Regions outside the DNA-binding domain are critical for proper in vivo specificity of an archetypal zinc finger transcription factor. *Nucleic Acids Res*. 2014;42(1):276–289.
55. 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.
56. Kadri Z, 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.
57. Goldfarb AN. Megakaryocytic programming by a transcriptional regulatory loop: A circle connecting RUNX1, GATA-1, and P-TEFb. *J Cell Biochem*. 2009;107(3):377–382.
58. Walkley CR, Sankaran VG, Orkin SH. Rb and hematopoiesis: stem cells to anemia. *Cell Div*. 2008;3:13.
59. Van Hook JW, Gill P, Cyr D, Kapur RP. Diamond-Blackfan anemia as an unusual cause of nonimmune hydrops fetalis: a case report. *J Reprod Med*. 1995;40(12):850–854.
60. Dunbar AE 3rd, Moore SL, Hinson RM. Fetal Diamond-Blackfan anemia associated with hydrops fetalis. *Am J Perinatol*. 2003;20(7):391–394.
61. Chou ST, et al. Trisomy 21 enhances human fetal erythro-megakaryocytic development. *Blood*. 2008;112(12):4503–4506.
62. Tunstall-Pedoe O, et al. Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations. *Blood*. 2008;112(12):4507–4511.
63. Roy A, et al. Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proc Natl Acad Sci U S A*. 2012;109(43):17579–17584.
64. Maclean GA, et al. Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells. *Proc Natl Acad Sci U S A*. 2012;109(43):17567–17572.
65. Klusmann JH, et al. Developmental stage-specific interplay of GATA1 and IGF signaling in fetal megakaryopoiesis and leukemogenesis. *Genes Dev*. 2010;24(15):1659–1672.
66. Toki T, et al. Naturally occurring oncogenic GATA1 mutants with internal deletions in transient abnormal myelopoiesis in Down syndrome. *Blood*. 2013;121(16):3181–3184.
67. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31(4):e15.
68. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. 2010;26(19):2363–2367.
69. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.
70. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy — analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307–315.
71. Letting DL, Rakowski C, Weiss MJ, Blobel GA. Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Mol Cell Biol*. 2003;23(4):1334–1340.
72. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3.
73. Benjamini Y, Hochberg Y. Controlling the false discovery rate — a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Methodol*. 1995;57(1):289–300.