SUPPLEMENTAL FIGURES

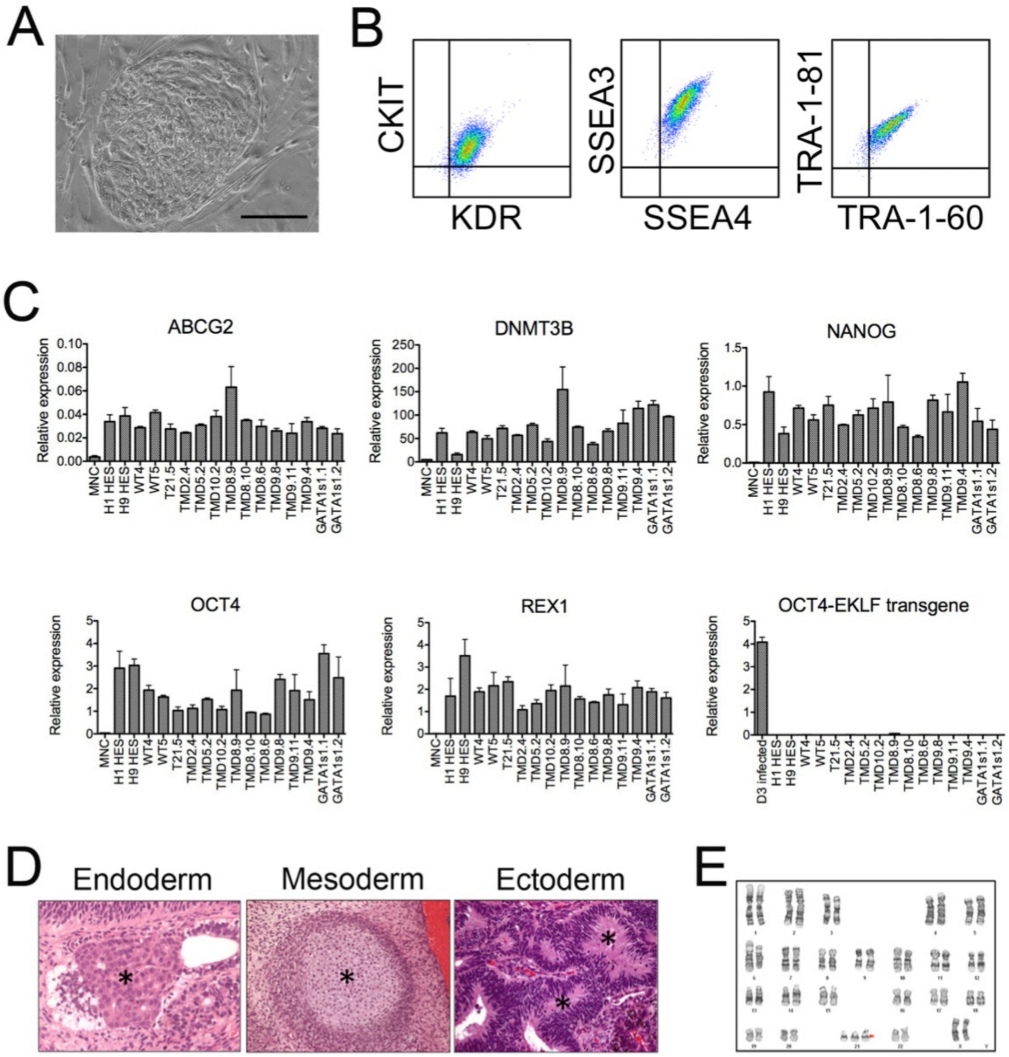


Figure S1. Generation of induced pluripotent stem cells from patient-derived samples. A. Photomicrograph demonstrating morphology of representative induced pluripotent stem cell (iPSC) colony. Scale bar, 200 μm. **B.** Representative flow cytometric analysis of iPSCs for cell surface expression of typical human pluripotent stem cell markers CKIT, KDR (VEGFR), SSEA3, SSEA4, TRA-1-81, and TRA-1-60. **C.** Expression levels of endogenous factors quantified by semi-quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) shown relative to cyclophilin levels. Bottom right panel shows expression levels of transgenic factors determined by qRT-PCR relative to cyclophilin levels. Results are shown as mean values plus or minus standard deviation (n = 3 replicates). MNC = mononuclear cells, HES = human embryonic stem cell. TMD = transient myeloproliferative disease. **D.** Representative photomicrographs of teratomas generated from iPSCs injected into (NOD/SCID) beige mice contain all three germ layers denoted by *. **E.** Representative karyotype analysis of iPSC clones showing trisomy 21.

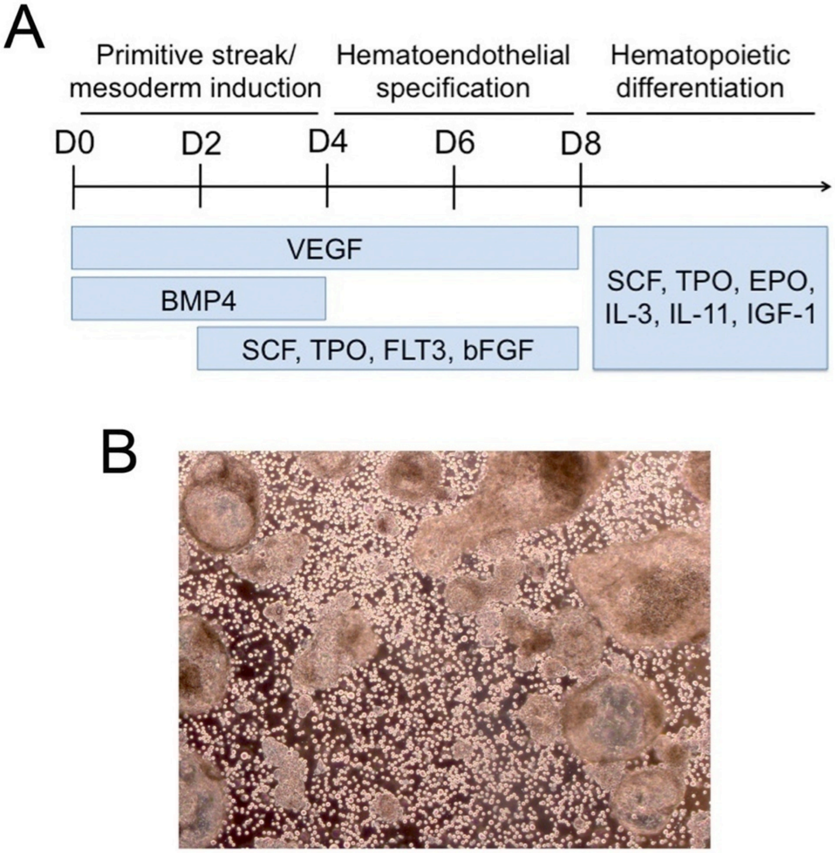


Figure S2. Hematopoietic differentiation via embryoid body (EB) formation. A. EBs were cultured in sequential cytokine combinations as indicated. VEGF, vascular endothelial growth factor; BMP4, bone morphogenic protein 4; SCF, stem cell factor; TPO, thrombopoietin; FLT3, FLT3-ligand; bFGF, fibroblast growth factor; EPO, erythropoietin; IL-3, interleukin-3; IL-11, interleukin-11; IGF-1, insulin growth factor-1. **B.** Photomicrograph of iPSC-derived EB culture with hematopoietic cells released into the medium. Original magnification, 4x.

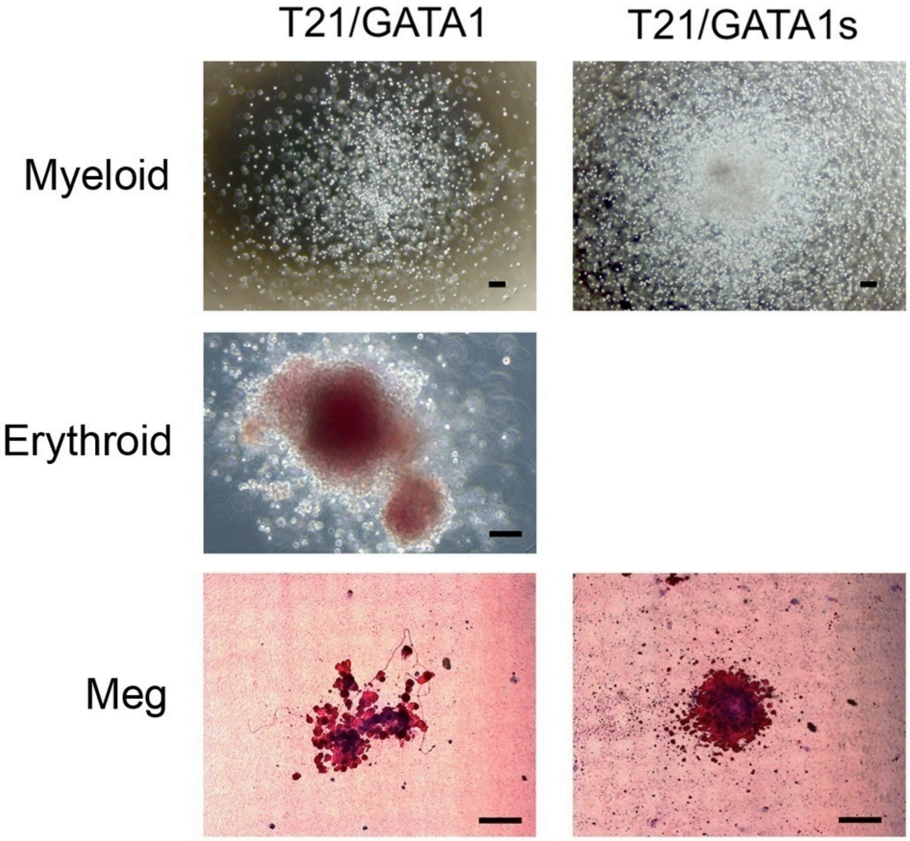


Figure S3. Hematopoietic colonies generated from induced pluripotent stem cell (iPSC)-derived progenitors. Representative myeloid, erythroid, and megakaryocyte (meg) colonies generated from T21/*wtGATA1* or T21/*GATA1s* iPSC-derived hematopoietic progenitors harvested on days 7-8 of embryoid body differentiation, and seeded into methylcellulose with EPO, SCF, IL3 and GMCSF, or Megacult collagen based assays with TPO, IL6 and IL3 for megakaryocyte colonies. Scale bars, 100 μm.

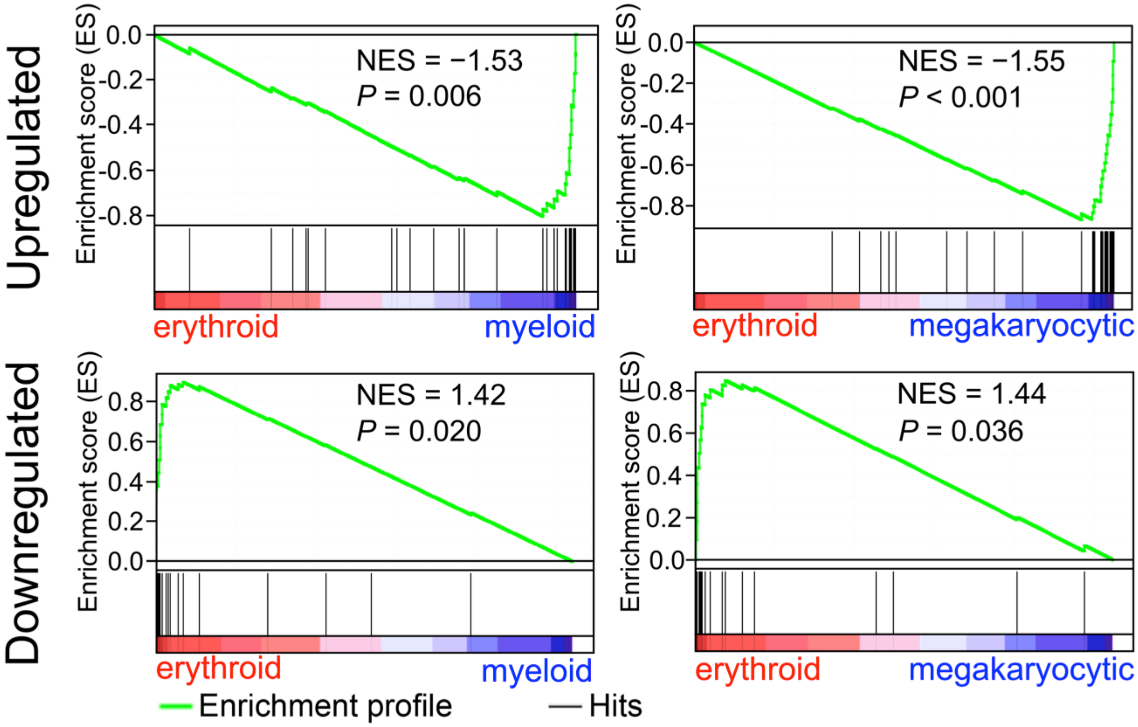


Figure S4. Gene set enrichment analysis on 56 differentially expressed genes (BH-FDR < 0.1) that are ≥ 2-fold up- or downregulated in T21/*GATA1s* as compared to T21/*wtGATA1.* Top panels show enrichment of 34 upregulated genes in T21/*GATA1s* as compared to T21/*wtGATA1* progenitors in a myeloid versus erythroid signature (top left), as well as in a megakaryocytic versus erythroid signature (top right). Bottom panels show enrichment of 22 downregulated genes in erythroid versus myeloid signature (bottom left), as well as in erythroid versus megakaryocytic signature (bottom right). NES, normalized enrichment score; *P* values shown are from modified Kolmogorov-Smirnov test as implemented in GSEA.

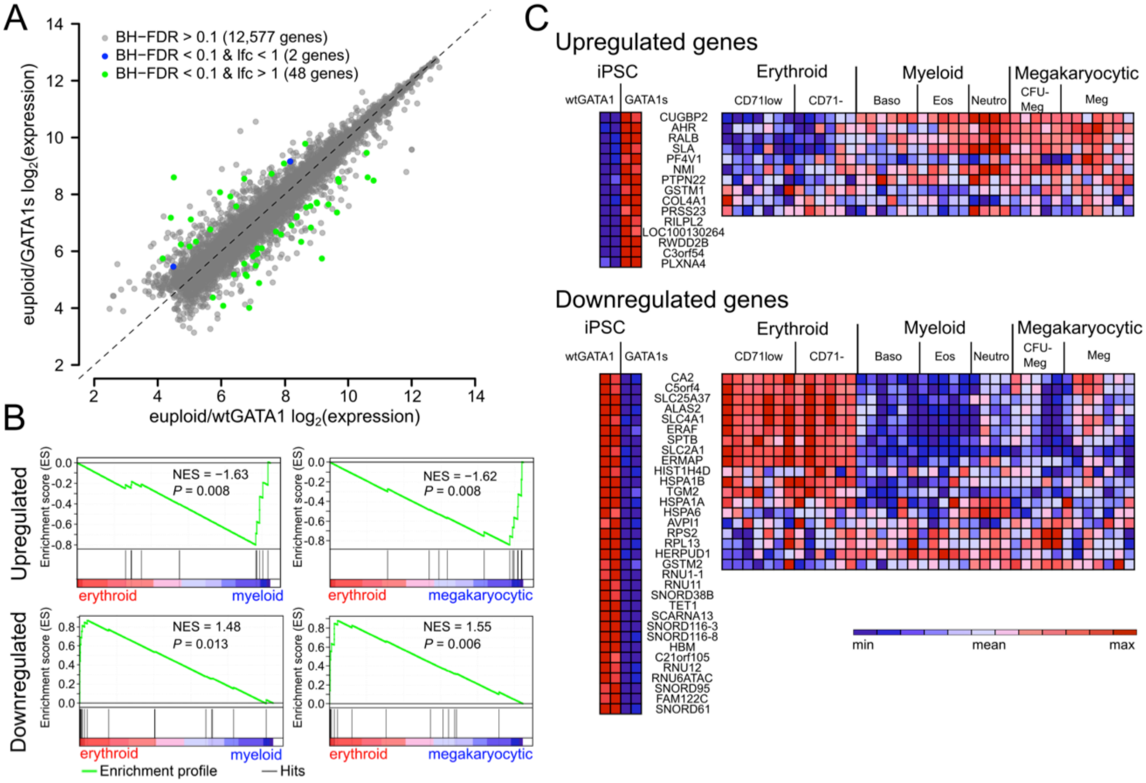
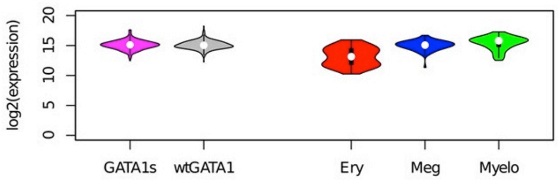
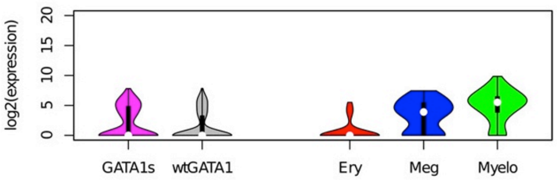


Figure S5. GATA1s downregulates an erythroid program and upregulates a myelo-megakaryocytic program in a genomewide transcriptome analysis in euploid iPSC-derived progenitors. A. Mean expression values of 12,627 expressed genes in euploid/*GATA1s* versus euploid/*wtGATA1* (2 replicates each) iPSC-derived CD43$^+$41$^+$235$^+$ progenitors. 50 genes were differentially expressed with a fold change of mean expression < 2 (2 genes, blue) or ≥ 2-fold (48 genes, green) between euploid/*GATA1s* and euploid/*wtGATA1*. **B.** GSEA showing enrichment of upregulated genes (top) in euploid/*GATA1s* as compared to euploid/*wtGATA1* progenitors in a myeloid versus erythroid signature, as well as in megakaryocytic versus erythroid signature, and enrichment of downregulated genes (bottom) in erythroid versus myeloid signature, as well as in erythroid versus megakaryocytic signature. NES, normalized enrichment score; *P* values shown are from modified Kolmogorov-Smirnov test as implemented in GSEA. **C.** Heat maps showing expression levels of upregulated (top) and downregulated (bottom) genes in euploid/*GATA1s* versus euploid/*wtGATA1* progenitors (top left and bottom left), as well as lineage-committed cells (top and bottom right) based on expressions levels in erythroid (7 replicates of CD34$^-$71^{low}GlyA$^+$, 6 replicates of CD34$^-$71$^-$GlyA$^+$ cells), myeloid (6 replicates of basophils, 5 replicates of eosinophils, 4 replicates of neutrophils), and megakaryocytic cells (5 replicates of CFU-megakaryocytes, CD34$^+$41$^+$61$^+$45-, 7 replicates of mature megakaryocytes, CD34$^-$41$^+$61$^+$45$^-$) from Novershtern *et al* (1). Color scheme is row normalized from blue to red corresponding to minimum to maximum expression values in a given row, respectively. Genes with no expression information in lineage-committed cells were not represented on microarrays from Novershtern *et al* (1).

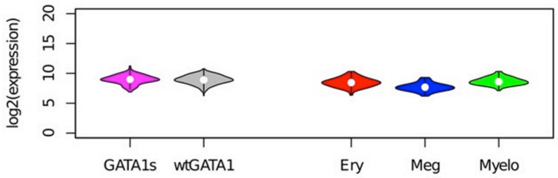
ACTB



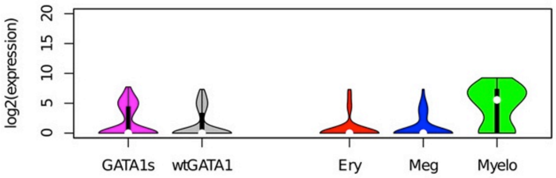
ARHGAP15*



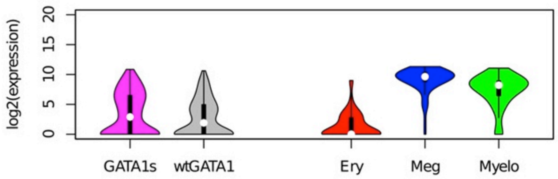
BACH1



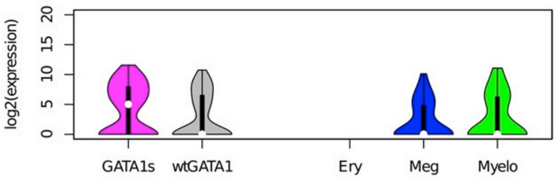
BCL11A*



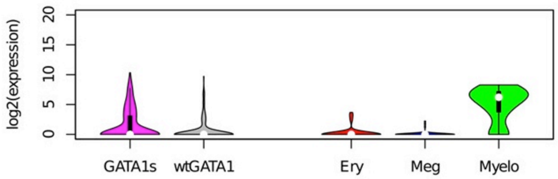
BIN2



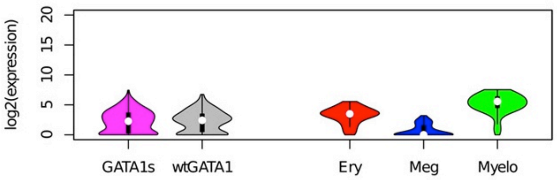
CD34**



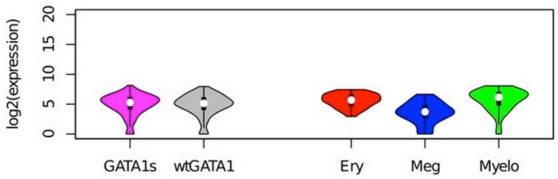
CEBPA**



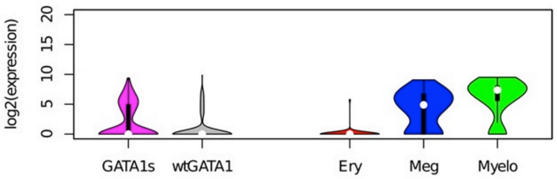
CEBPB



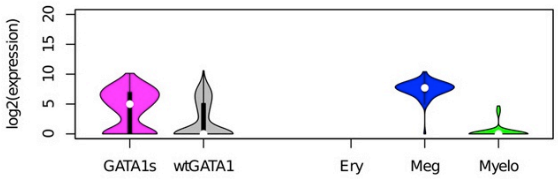
CEBPG



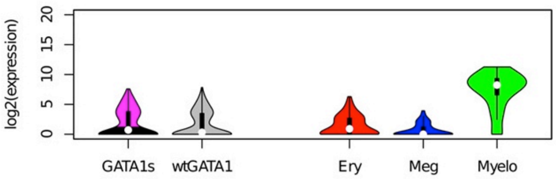
CFH**



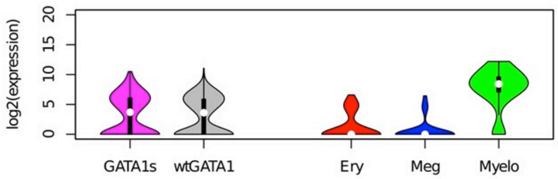
COL24A1**



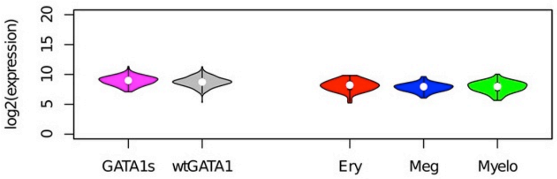
CSF2RA

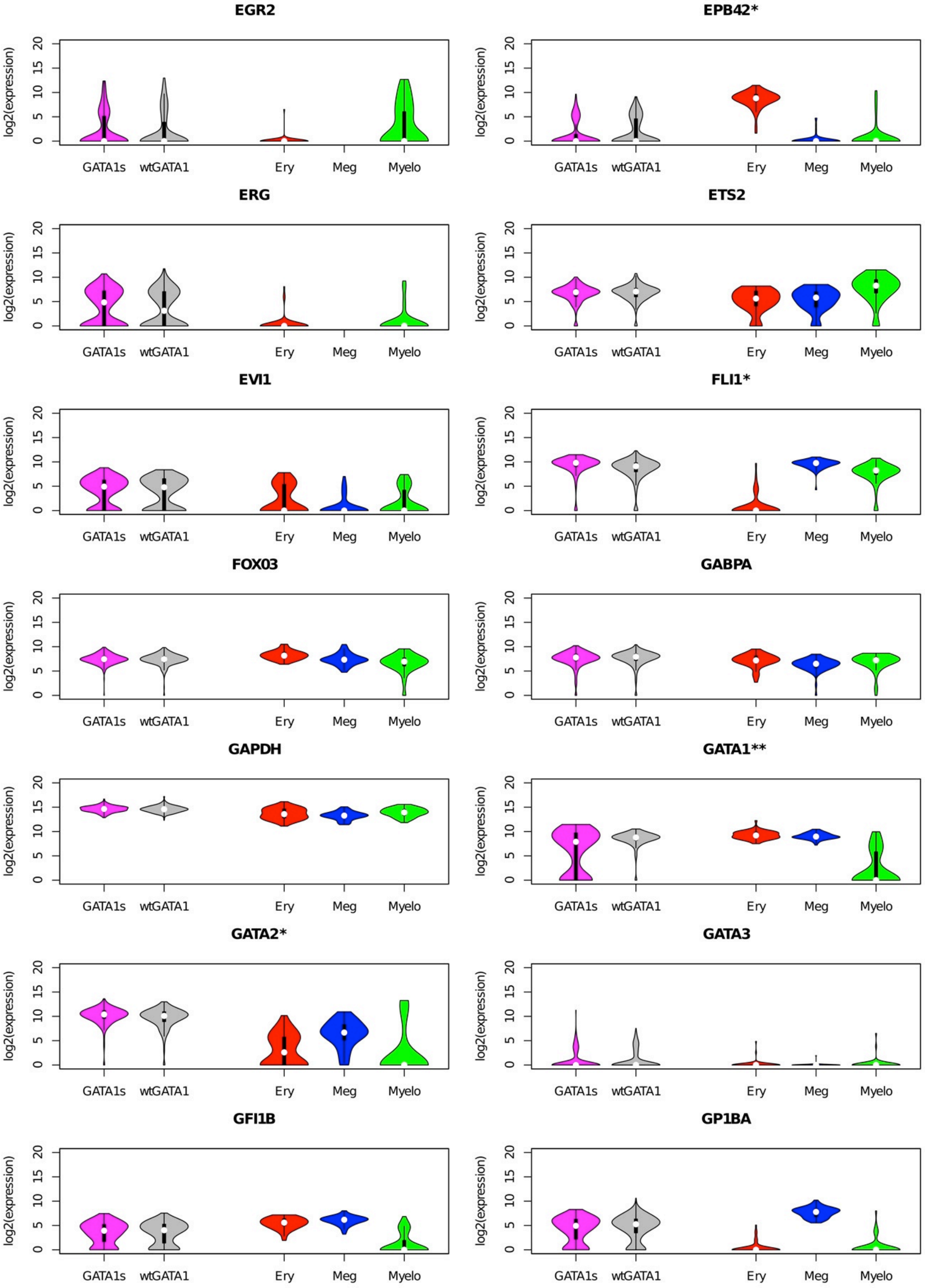


CSF3R

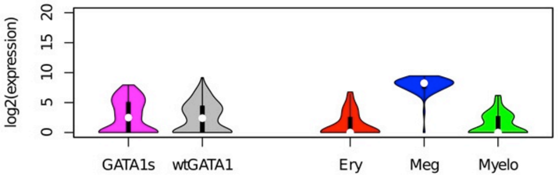


DYRK1A*

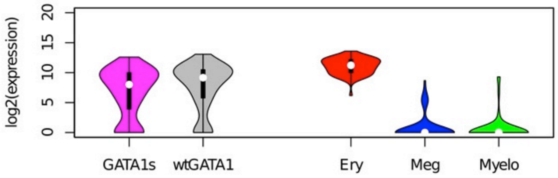



EGR2
EPB42*
ERG
ETS2
EVI1
FLI1*
FOX03
GABPA
GAPDH
GATA1**
GATA2*
GATA3
GFI1B
GP1BA
log2(expression)
0
5
10
15
20
GATA1s
wtGATA1
Ery
Meg
Myelo

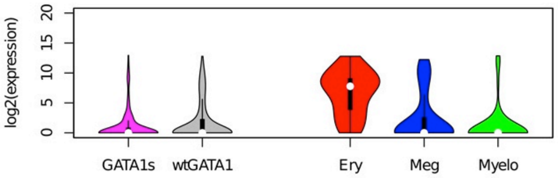
GP9



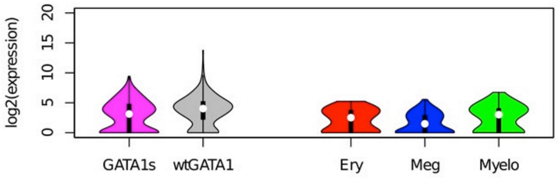
GYPA*



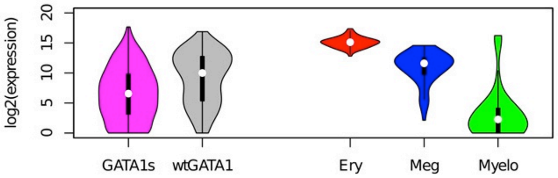
HBA2*



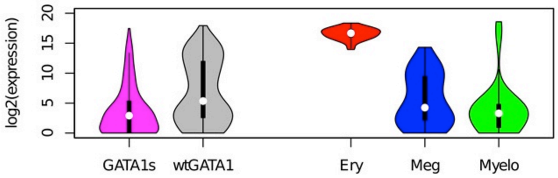
HBB*



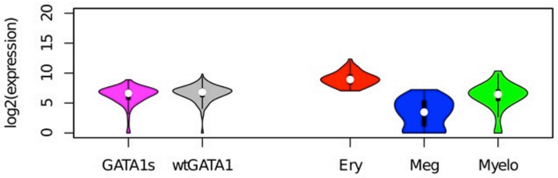
HBG1**



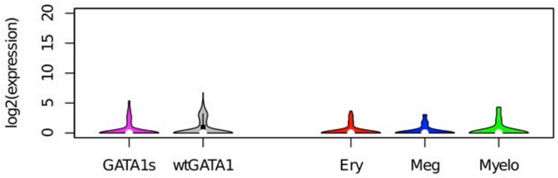
HBZ**



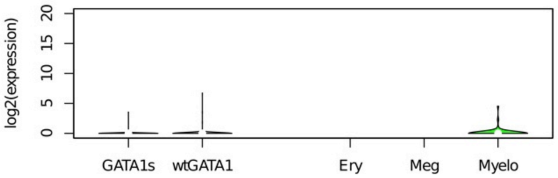
HMBS*



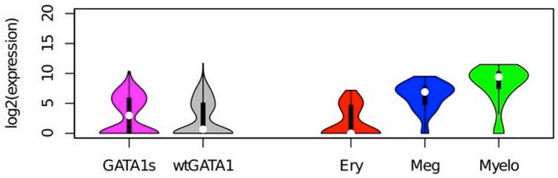
HOXA9



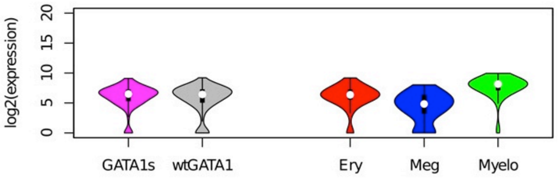
HOXA10



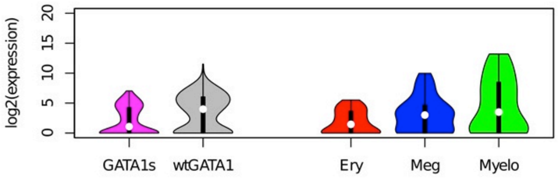
IFI16



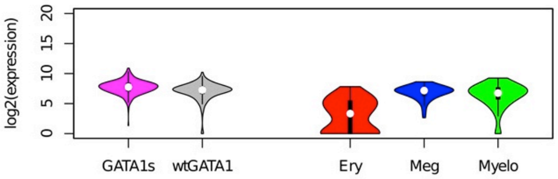
IKZF1



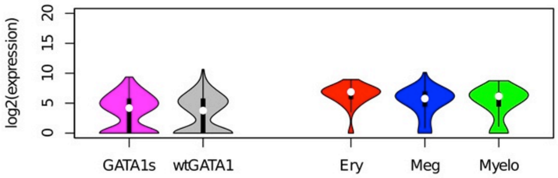
IL8**

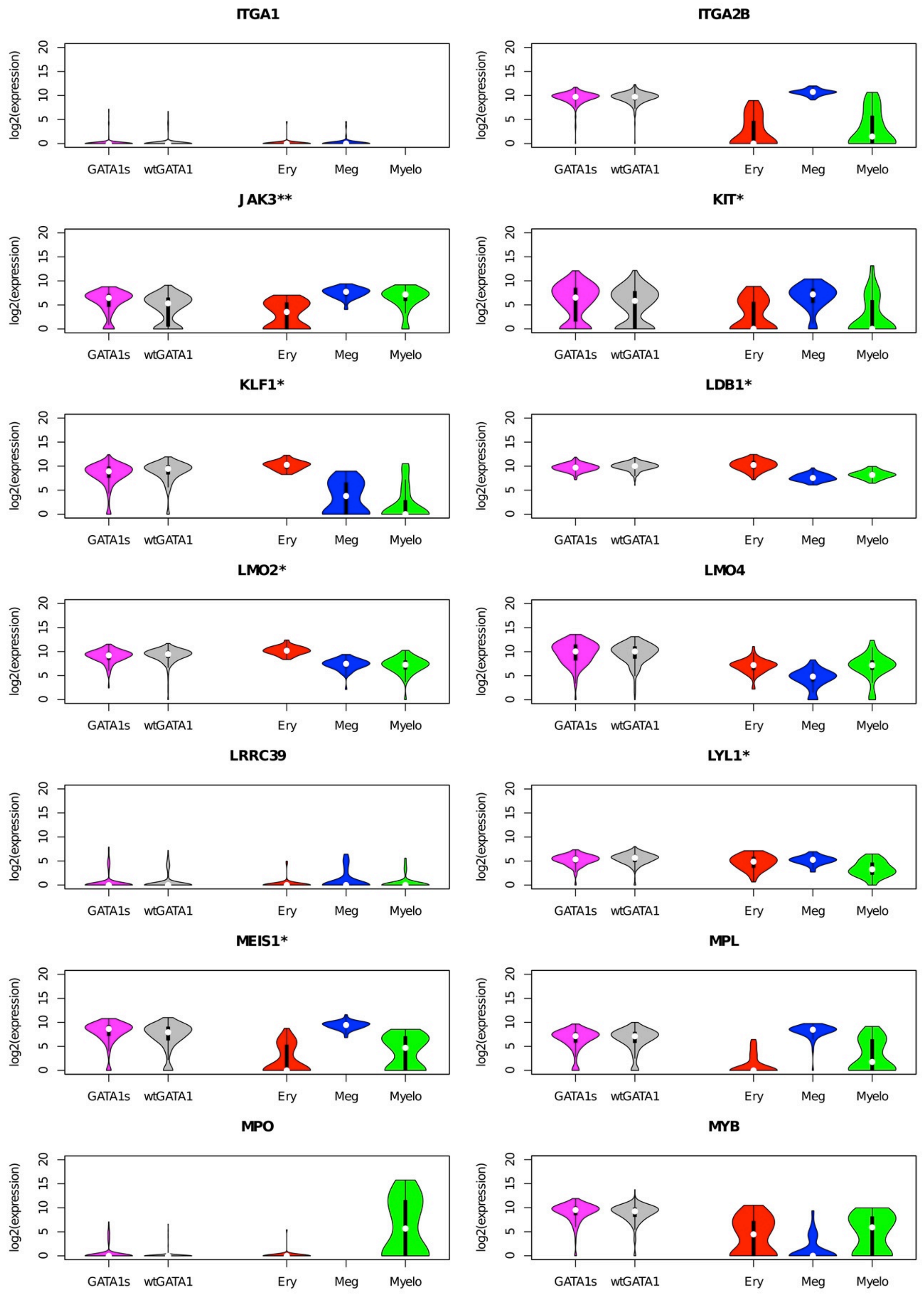


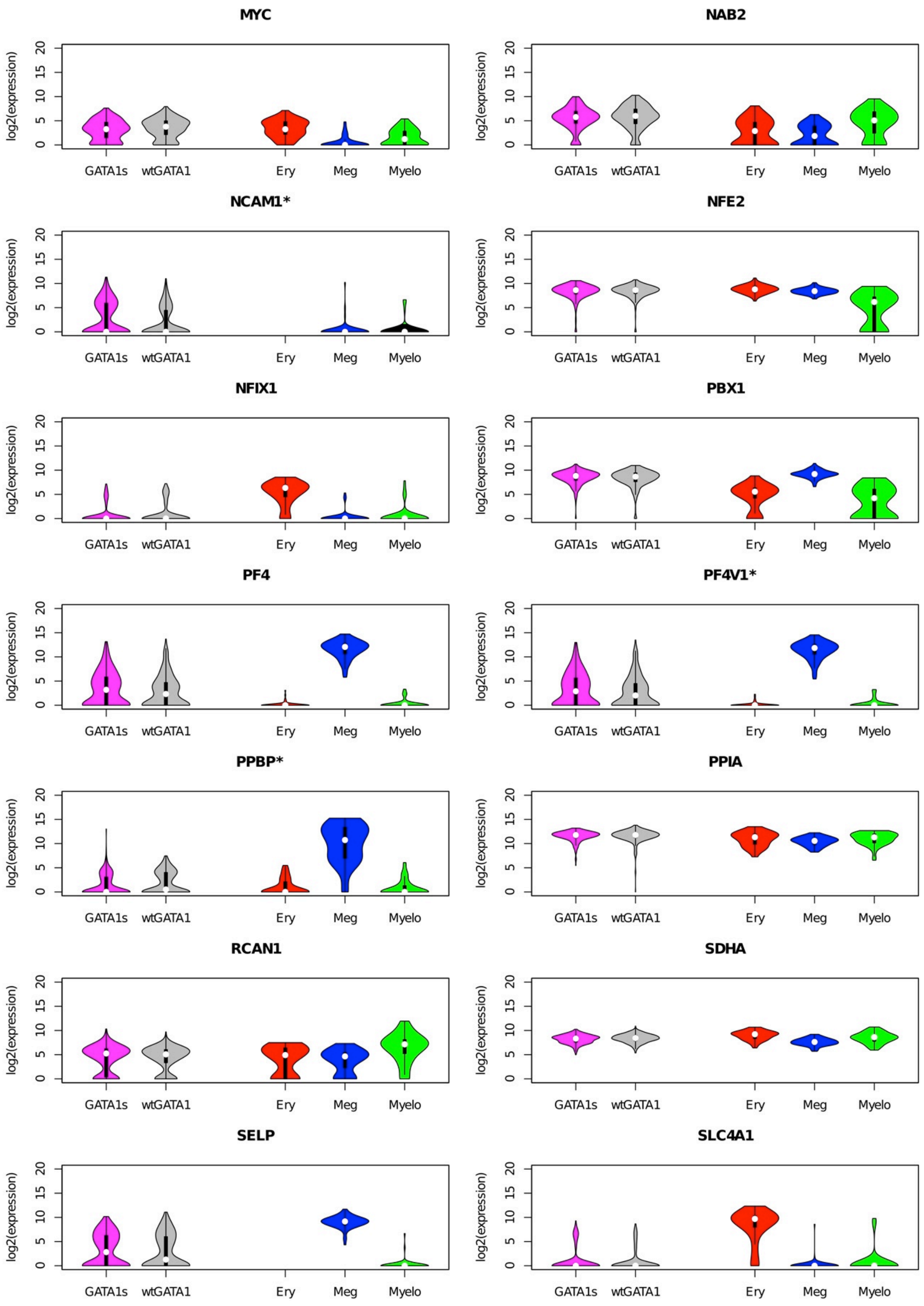
INF2*

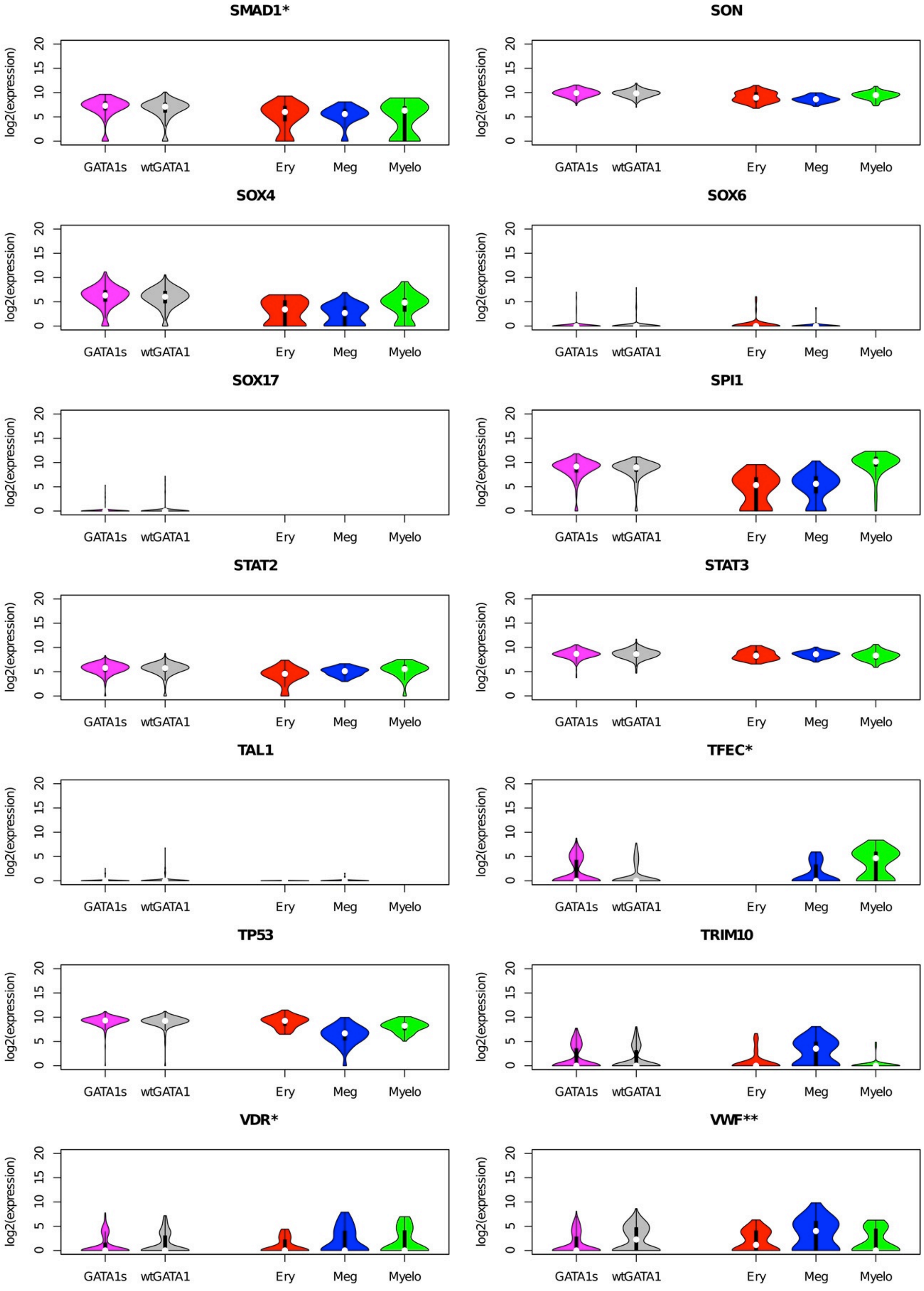


IRF1




ITGA1
ITGA2B
JAK3**
KIT*
KLF1*
LDB1*
LMO2*
LMO4
LRRC39
LYL1*
MEIS1*
MPL
MPO
MYB
log2(expression)
0
5
10
15
20
GATA1s
wtGATA1
Ery
Meg
Myelo


MYC
NAB2
NCAM1*
NFE2
NFIX1
PBX1
PF4
PF4V1*
PPBP*
PPIA
RCAN1
SDHA
SELP
SLC4A1
log2(expression)
0
5
10
15
20
GATA1s
wtGATA1
Ery
Meg
Myelo


SMAD1*
SON
SOX4
SOX6
SOX17
SPI1
STAT2
STAT3
TAL1
TFEC*
TP53
TRIM10
VDR*
VWF**
log2(expression)
0
5
10
15
20
GATA1s
wtGATA1
Ery
Meg
Myelo

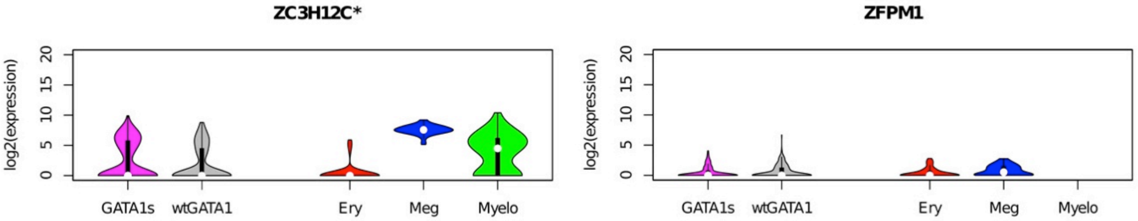


Figure S6. Violin plots showing distributions of single cell expression levels for each of the analyzed genes in *GATA1s* and *wtGATA1* progenitors, and lineage-committed erythroid, megakaryocytic, and myeloid cells. Numbers of cells whose expression values for a given gene were included in each violin: 274 GATA1s- and 311 wtGATA1-expressing iPSC-derived progenitors, and 57 erythroid, 61 megakaryocytic, and 52 myeloid iPSC-derived lineage-committed cells. Single asterisk (*) next to a gene symbol marks genes that are differentially expressed between *GATA1s* and *wtGATA1* progenitors (FDR < 0.05; Mann-Whitney *U* test followed by BH-FDR correction). Double asterisk (**) marks differentially expressed genes (FDR < 0.05) that are ≥ 2-fold up- or downregulated in *GATA1s* as compared to *wtGATA1* progenitors. lfc – $\log_2$ of fold change of mean gene expression between *GATA1s* and *wtGATA1* progenitors. Violin plots are organized in an alphabetical order. Violin plots for *F10* and *WNT10A* are not shown, because these genes displayed no detectable expression across all cell types analyzed. Violin plots for *CSF1R, GFI1, RUNX1, HBE1, ALAS2*, and *EPOR* are in Figure 5D.

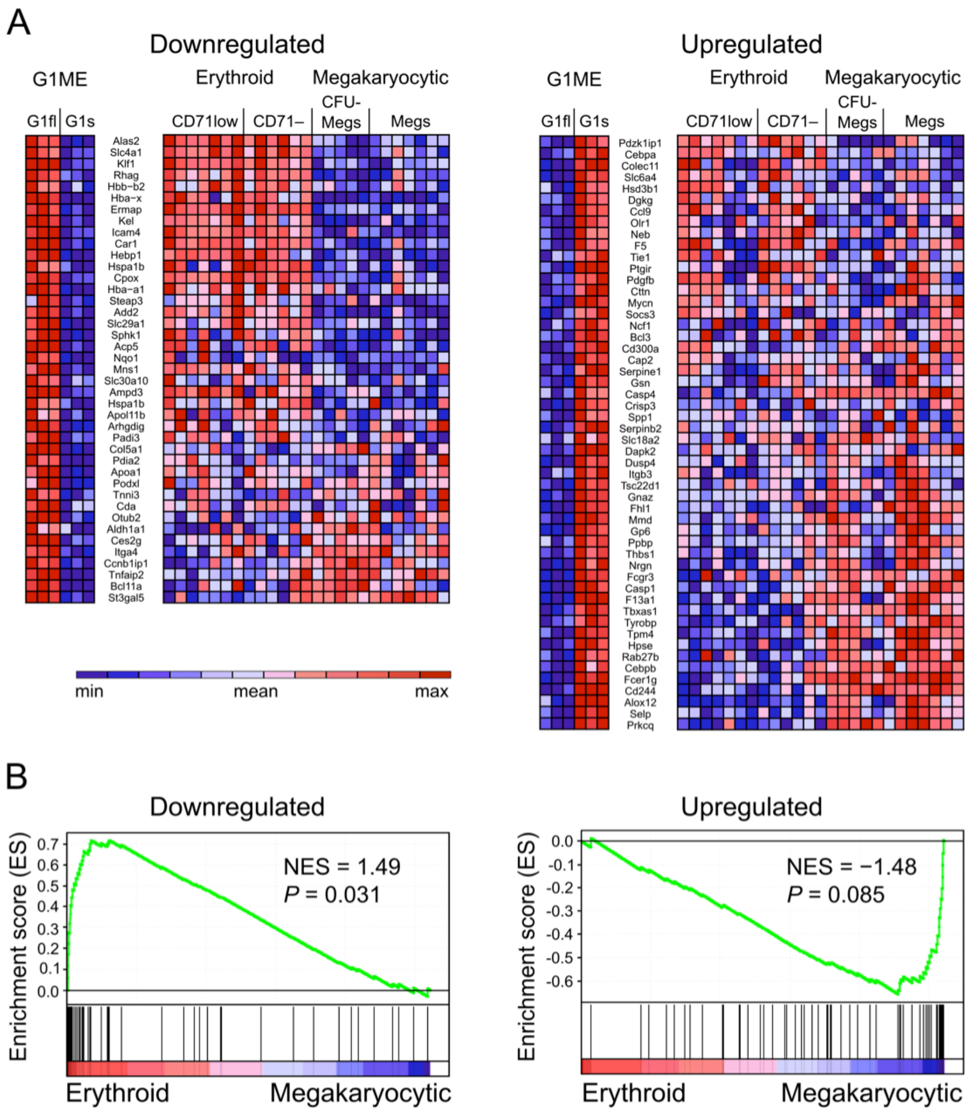


Figure S7. GATA1s downregulates an erythroid transcriptional program and upregulates a megakaryocytic program in *Gata1$^-$* murine megakaryocyte-erythroid progenitors at 42 hours post-transduction. A. Heat maps showing expression levels of genes downregulated (first from the left) and upregulated (third from the left) in G1ME/*GATA1s* vs. G1ME/*GATA1fl* (3 replicates each; FDR < 0.1, > 2-fold change in expression). Second and fourth heat maps from the left show expression levels of downregulated and upregulated genes, respectively, in human erythroid (7 replicates of CD34$^-$71^{low}GlyA$^+$, 6 replicates of CD34$^-$71$^-$GlyA$^+$ cells) and megakaryocytic cells (5 replicates of CFU-megakaryocytes, CD34$^+$41$^+$61$^+$45-, 7 replicates of mature megakaryocytes, CD34$^-$41$^+$61$^+$45$^-$) from Novershtern *et al* (1). Color scheme is row normalized from blue to red corresponding to minimum to maximum expression values in a given row, respectively. **B.** GSEA showing enrichment of genes that are downregulated by GATA1s (left) for erythroid as compared to megakaryocytic genes, and enrichment of genes that are upregulated by GATA1s (right) for megakaryocytic as compared to erythroid genes.

NES, normalized enrichment score; *P* values shown are from modified Kolmogorov-Smirnov test as implemented in GSEA. Since human erythroid and megakaryocytic expression data sets were used as gene signatures in panels A and B, only those down- and upregulated genes that have orthologs in the human genome were listed in heat maps and included in the GSEA.

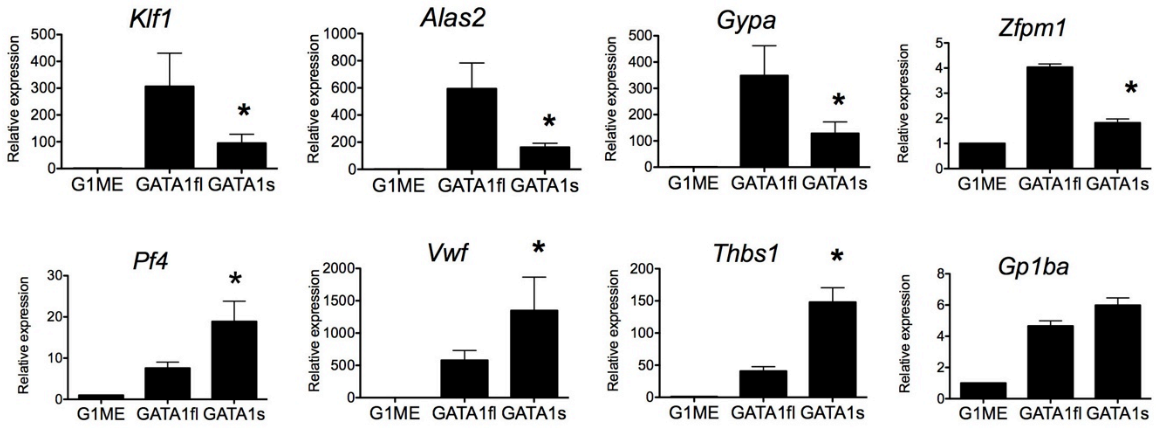


Figure S8. Expression of selected gene targets in G1ME cells transduced with GATA1fl or GATA1s. Average expression of selected erythroid (top) and megakaryocytic (bottom) GATA1 target genes 42 hours post-transduction +/- SD (n = 4 independent experiments). **P* < 0.05 (2-tailed Student's *t* test).

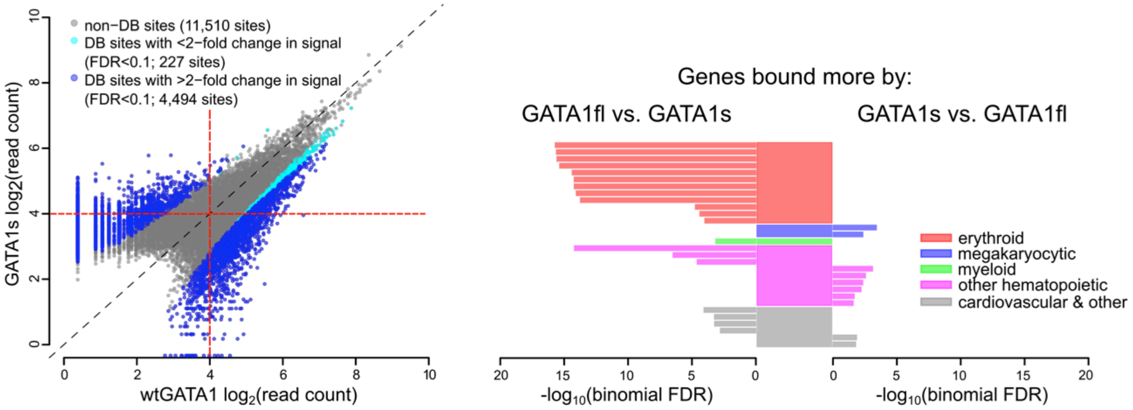


Figure S9. Sites bound more by GATA1s display lower binding signal and less significant functional enrichment as compared to sites bound more by GATA1fl. Left, as described in Figure 7D. Red dashed lines represent a threshold of binding signal, applied for the analysis on the right, separating sites with lower signal from sites with higher signal. **Right,** Functional enrichment analysis using GREAT performed on differentially bound sites with > 2-fold change in binding signal and normalized read count of > 4. Plotted are significance values for top 10 “mouse phenotype” and “GO biological process” enrichment terms (30 terms total; there were no significant “GO biological process” terms for genes bound more by GATA1s vs. GATA1fl at sites with higher binding signal), classified as erythroid, megakaryocytic, myeloid, other hematopoietic, or cardiovascular and other.

SUPPLEMENTAL TABLES

Table S1. Induced pluripotent stem cell lines used in this study. WT, wild type; T21, trisomy 21; TMD, transient myeloproliferative disorder; PB, peripheral blood; MNC, mononuclear cells; Retro, pMXs-based retroviruses, Lenti, hSTEMMCA-loxP lentivirus; OSKM, *OCT4*, *SOX2*, *KLF4*, *MYC*. #Lines characterized in (2). *Line purchased from George Daley lab, **NCBI reference sequence NM_002049, nucleotide 1 = first nucleotide of exon 1. Rows in different shades of gray represent isogenic lines derived from the same patient with and without *GATA1* mutations.

Name	**Cell of Origin**	**Reprogramming Vector**	**Karyotype**	***GATA1*** **
WT1*	Neonatal fibroblast	Retro: OSKM	46, XY	WT
WT2#	Fetal stromal cell	Retro: OSKM	46, XY	WT
WT3#	Fetal MNC	Lenti: OSKM	46, XY	WT
WT4	Fetal MNC	Lenti: OSKM	46, XY	WT
WT5	Fetal MNC	Lenti: OSKM	46, XX	WT
T21.1#	Neonatal fibroblast	Retro: OSKM	47, XY, +21	WT
T21.2#	Fetal stromal cell	Retro: OSKM	47, XY, +21	WT
T21.3#	Fetal stromal cell	Retro: OSKM	47, XY, +21	WT
T21.4#	Fetal MNC	Lenti: OSKM	47, XY, +21	WT
T21.5	Fetal MNC	Lenti: OSKM	47, XY, +21	WT
TMD2.4	PB MNC	Lenti: OSKM	47, XY, +21	g.4605del
TMD5.2	PB MNC	Lenti: OSKM	47, XY, +21	g.4757G>A
TMD10.2	PB MNC	Lenti: OSKM	47, XX, +21	g.4703dup
TMD8.9	PB MNC	Lenti: OSKM	47, XX, +21	g.4652G>T
TMD8.10	PB MNC	Lenti: OSKM	47, XX, +21	WT
TMD8.6	PB MNC	Lenti: OSKM	47, XX, +21	WT
TMD9.8	PB MNC	Lenti: OSKM	47, XX, +21	g.4500del_ins
TMD9.11	PB MNC	Lenti: OSKM	47, XX, +21	g.4500del_ins
TMD9.4	PB MNC	Lenti: OSKM	47, XX, +21	WT
GATA1s1.1	Adult fibroblasts	Lenti: OSKM	46, XY	g.4755G>C
GATA1s1.2	Adult fibroblasts	Lenti: OSKM	46, XY	g.4755G>C

Table S2. Genes differentially expressed between T21/*GATA1s* and T21/*wtGATA1* iPSC-derived progenitors (BH-FDR < 0.1; lfc ≥ 1) identified using moderated *t* test ("limma" package in R). Probability, probability that a gene is differentially expressed (from "limma" package in R); lfc, $\log_2$ of fold change (negative numbers correspond to downregulation in GATA1s- as compared to wtGATA1-expressing cells, whereas positive numbers correspond to upregulation in GATA1s- as compared to wtGATA1-expressing cells); GATA1 targets, genes that are bound by GATA1 in human PBDE and/or PBDEFetal cells at one or more sites within a 10kb gene neighborhood, i.e. 10kb upstream of TSS + gene body + 10kb downstream of TES ("1" – GATA1 target, "0" – not a GATA1 target). Genes are ordered from largest to smallest absolute value of lfc.

Downregulated in T21/*GATA1s* vs. T21/*wtGATA1*					
Gene symbol	***P* value**	**BH-FDR**	**Probability**	**lfc**	**GATA1 targets**
HBZ	2.90E-06	3.94E-03	0.99	-3.96	1
AHSP	7.94E-04	5.79E-02	0.45	-2.42	1
RELN	7.44E-06	5.78E-03	0.98	-2.04	1
ALAS2	4.74E-06	4.80E-03	0.99	-2.01	1
SPTA1	3.02E-04	3.53E-02	0.67	-2.01	1
SLC30A10	3.79E-07	1.94E-03	1.00	-1.81	1
APOC1	2.72E-06	3.94E-03	0.99	-1.61	1
HBA1	1.15E-03	7.11E-02	0.37	-1.60	1
HBA2	1.15E-03	7.11E-02	0.37	-1.60	1
MYH10	2.39E-05	1.28E-02	0.95	-1.57	1
HBG1	5.27E-04	4.69E-02	0.55	-1.48	1
SLC25A21	1.72E-03	8.68E-02	0.28	-1.44	1
SLC25A37	5.34E-07	1.94E-03	1.00	-1.34	1
OCIAD2	8.37E-07	2.28E-03	1.00	-1.18	0
NEDD4L	3.36E-05	1.30E-02	0.94	-1.14	1
LY6G6D	1.17E-04	2.16E-02	0.83	-1.13	1
HBE1	2.38E-04	3.05E-02	0.72	-1.09	1
BLVRB	3.20E-05	1.29E-02	0.94	-1.08	1
GSTA1	2.39E-03	9.68E-02	0.22	-1.05	0
ANKRD26	1.17E-04	2.16E-02	0.83	-1.03	0
KEL	6.19E-05	1.72E-02	0.90	-1.03	1
JHDM1D	1.40E-04	2.34E-02	0.81	-1.01	1
Upregulated in T21/*GATA1s* vs. T21/*wtGATA1*					
Gene symbol	***P* value**	**BH-FDR**	**Probability**	**lfc**	**GATA1 targets**
IFI16	4.43E-04	4.30E-02	0.59	2.27	1

PF4V1	1.13E-03	7.11E-02	0.37	2.03	0
CFH	1.30E-03	7.57E-02	0.34	2.01	1
PARP9	1.79E-03	8.81E-02	0.27	2.01	0
IFIT1	1.92E-03	8.97E-02	0.26	1.94	1
ARHGAP15	2.02E-04	2.78E-02	0.75	1.87	1
TFEC	6.48E-05	1.72E-02	0.89	1.79	0
ZC3H12C	4.75E-05	1.61E-02	0.92	1.72	0
NCAM1	2.82E-06	3.94E-03	0.99	1.68	0
BIN2	1.46E-04	2.41E-02	0.80	1.63	1
COL24A1	6.67E-04	5.32E-02	0.49	1.62	0
IL8	1.36E-06	2.96E-03	1.00	1.49	1
P2RY13	2.18E-04	2.96E-02	0.73	1.41	0
CD180	6.03E-06	5.46E-03	0.99	1.39	1
PDE3A	1.05E-03	6.93E-02	0.39	1.38	1
P2RY14	1.62E-03	8.47E-02	0.29	1.38	1
RGS18	7.96E-05	1.84E-02	0.87	1.35	0
BIRC3	2.44E-04	3.05E-02	0.71	1.33	1
GPR171	2.06E-03	9.12E-02	0.25	1.32	1
S100B	6.30E-04	5.16E-02	0.51	1.29	0
ATP8B4	1.54E-04	2.49E-02	0.79	1.26	1
LPAR4	6.41E-04	5.16E-02	0.50	1.26	0
MIR221	1.26E-04	2.25E-02	0.82	1.25	0
ABCB1	1.83E-04	2.59E-02	0.76	1.24	1
CD44	6.04E-05	1.72E-02	0.90	1.22	1
CXCL2	8.78E-04	6.08E-02	0.43	1.22	1
CXCL6	8.57E-05	1.94E-02	0.87	1.19	0
P2RY12	1.34E-03	7.69E-02	0.33	1.13	1
FCGR2A	5.04E-05	1.66E-02	0.91	1.13	1
RGS1	7.85E-05	1.84E-02	0.88	1.11	0
MMRN1	7.28E-05	1.76E-02	0.88	1.09	1
FYB	7.37E-04	5.60E-02	0.47	1.04	1
FUT8	6.27E-04	5.16E-02	0.51	1.03	1
EGF	1.35E-03	7.70E-02	0.33	1.01	0

Table S3. List of 94 selected genes whose expression was measured at a single cell level in iPSC-derived progenitors expressing wtGATA1 or GATA1s, as well as in iPSC-derived lineage-committed erythroid, megakaryocytic, or myeloid cells. Genes are listed in an alphabetical order. An asterisk next to a gene symbol marks housekeeping genes.

Gene symbol	Description
ACTB *	actin, beta
ALAS2	aminolevulinate, delta-, synthase 2
ARHGAP15	Rho GTPase activating protein 15
BACH1	BTB and CNC homology 1, basic leucine zipper transcription factor 1
BCL11A	B-cell CLL/lymphoma 11A (zinc finger protein)
BIN2	bridging integrator 2
CD34	CD34 molecule
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
CEBPG	CCAAT/enhancer binding protein (C/EBP), gamma
CFH	complement factor H
COL24A1	collagen, type XXIV, alpha 1
CSF1R	colony stimulating factor 1 receptor
CSF2RA	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)
CSF3R	colony stimulating factor 3 receptor (granulocyte)
DYRK1A	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A
EGR2	early growth response 2
EPB42	erythrocyte membrane protein band 4.2
EPOR	erythropoietin receptor
ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
EVI1	ecotropic viral integration site-1
F10	coagulation factor X
FLI1	friend leukemia virus integration 1
FOX03	forkhead box 03
GABPA	GA binding protein transcription factor, alpha subunit 60kDa
GAPDH *	glyceraldehyde-3-phosphate dehydrogenase
GATA1	GATA binding protein 1 (globin transcription factor 1)
GATA2	GATA binding protein 2
GATA3	GATA binding protein 3
GFI1	growth factor independent 1 transcription repressor
GFI1B	growth factor independent 1B transcription repressor
GP1BA	glycoprotein Ib (platelet), alpha polypeptide
GP9	glycoprotein IX (platelet)
GYPA	glycophorin A (MNS blood group)
HBA2	hemoglobin, alpha 2
HBB	hemoglobin, beta
HBE1	hemoglobin, epsilon 1
HBG1	hemoglobin, gamma A
HBZ	hemoglobin, zeta

HMBS	hydroxymethylbilane synthase
HOXA10	homeobox A10
HOXA9	homeobox A9
IFI16	interferon, gamma-inducible protein 16
IKZF1	IKAROS family zinc finger 1 (Ikaros)
IL8	interleukin 8
INF2	inverted formin, FH2 and WH2 domain containing
IRF1	interferon regulatory factor 1
ITGA1	integrin, alpha 1
ITGA2B	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)
JAK3	Janus kinase 3
KIT	proto-oncogene tyrosine-protein kinase Kit) (c-kit) (CD117 antigen)
KLF1	Kruppel-like factor 1 (erythroid)
LDB1	LIM domain binding 1
LMO2	LIM domain only 2 (rhombotin-like 1)
LMO4	LIM domain only 4
LRRC39	leucine rich repeat containing 39
LYL1	lymphoblastic leukemia derived sequence 1
MEIS1	Meis homeobox 1
MPL	myeloproliferative leukemia virus oncogene
MPO	myeloperoxidase
MYB	v-myb myeloblastosis viral oncogene homolog (avian)
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)
NCAM1	neural cell adhesion molecule 1
NFE2	nuclear factor (erythroid-derived 2), 45kDa
NFIX1	nuclear factor I/X type 1
PBX1	pre-B-cell leukemia homeobox 1
PF4	platelet factor 4
PF4V1	platelet factor 4 variant 1
PPBP	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)
PPIA	peptidylprolyl isomerase A (cyclophilin A)
RCAN1	regulator of calcineurin 1 (DSCR1)
RUNX1	runt-related transcription factor 1
SDHA *	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)
SELP	selectin P (granule membrane protein 140kDa, antigen CD62)
SLC4A1	solute carrier family 4, anion exchanger, member 1
SMAD1	SMAD family member 1
SON	SON DNA binding protein
SOX17	SRY (sex determining region Y)-box 17
SOX4	SRY (sex determining region Y)-box 4
SOX6	SRY (sex determining region Y)-box 6
SPI1	spleen focus forming virus (SFFV) proviral integration oncogene spi1
STAT2	signal transducer and activator of transcription 2, 113kDa
STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)
TAL1	T-cell acute lymphocytic leukemia 1
TFEC	transcription factor EC
TP53	tumor protein p53
TRIM10	tripartite motif-containing 10

VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor
VWF	von Willebrand factor
WNT10A	wingless-type MMTV integration site family, member 10A
ZC3H12C	zinc finger CCCH-type containing 12C
ZFPM1	zinc finger protein, FOG family member 1

Table S4. Forty differentially expressed genes among 94 genes assayed at a single cell level between GATA1s- and wtGATA1-expressing single iPSC-derived CD43$^+$41$^+$235$^+$ progenitors (BH-FDR < 0.05), identified using Mann-Whitney *U* test. Mean, mean log$_2$(expression); Median, median log$_2$(expression); lfc, log$_2$ of fold change (negative numbers correspond to downregulation in GATA1s- as compared to wtGATA1-expressing cells, whereas positive numbers correspond to upregulation in GATA1s- as compared to wtGATA1-expressing cells). Genes are ordered from largest to smallest absolute value of lfc of means.

Downregulated (on average) in *GATA1s* vs. *wtGATA1* progenitors								
Gene symbol	***P* value**	**BH-FDR**	**GATA1s**		**wtGATA1**		**lfc of means**	**lfc of medians**
			Mean	**Median**	**Mean**	**Median**		
HBZ	1.94E-13	6.22E-12	3.78	2.90	6.93	5.34	-3.15	-2.44
HBG1	1.28E-11	2.46E-10	6.54	6.57	9.08	10.00	-2.53	-3.44
GATA1	7.08E-06	4.25E-05	6.07	7.89	8.49	8.78	-2.43	-0.89
ALAS2	1.99E-11	2.73E-10	5.20	5.82	7.60	8.78	-2.40	-2.96
HBE1	1.90E-06	1.22E-05	4.80	3.13	7.13	7.03	-2.33	-3.90
EPOR	1.84E-17	8.82E-16	2.97	3.02	4.84	5.49	-1.86	-2.47
IL8	3.98E-10	4.25E-09	2.09	1.07	3.57	3.99	-1.48	-2.91
VWF	3.87E-08	2.86E-07	1.38	0.00	2.47	2.26	-1.09	-2.26
GYPA	8.50E-04	3.02E-03	6.74	8.01	7.70	9.16	-0.96	-1.15
HBB	6.87E-05	3.14E-04	2.89	3.13	3.74	4.03	-0.85	-0.90
HBA2	1.81E-03	5.80E-03	1.06	0.00	1.84	0.00	-0.78	0.00
EPB42	2.40E-03	7.45E-03	1.37	0.00	1.97	0.00	-0.60	0.00
KLF1	1.41E-04	5.40E-04	8.29	8.92	8.87	9.44	-0.58	-0.52
PPBP	1.44E-03	4.78E-03	1.47	0.00	1.99	0.51	-0.52	-0.51
VDR	1.46E-02	3.50E-02	1.04	0.00	1.38	0.00	-0.35	0.00
HMBS	1.01E-02	2.56E-02	6.25	6.64	6.59	6.83	-0.33	-0.19
LYL1	8.81E-04	3.02E-03	5.20	5.35	5.49	5.68	-0.29	-0.33
LDB1	5.02E-05	2.41E-04	9.70	9.70	9.93	10.02	-0.23	-0.32
LMO2	8.76E-03	2.34E-02	9.01	9.21	9.21	9.47	-0.20	-0.26
Upregulated (on average) in *GATA1s* vs. *wtGATA1* progenitors								
Gene symbol	***P* value**	**BH-FDR**	**GATA1s**		**wtGATA1**		**lfc of means**	**lfc of medians**
			Mean	**Median**	**Mean**	**Median**		
COL24A1	9.74E-13	2.34E-11	4.17	4.99	2.23	0.00	1.95	4.99
CSF1R	9.73E-08	6.67E-07	4.52	5.60	2.83	0.00	1.69	5.60
CFH	4.94E-11	5.92E-10	2.09	0.00	0.77	0.00	1.32	0.00
CD34	1.24E-04	4.96E-04	4.37	5.01	3.10	0.00	1.26	5.01
JAK3	1.04E-09	9.99E-09	5.59	6.47	4.34	5.33	1.26	1.13
CEBPA	3.43E-09	2.75E-08	1.66	0.00	0.60	0.00	1.06	0.00
GFI1	9.79E-05	4.27E-04	6.01	7.57	4.96	6.53	1.04	1.05
NCAM1	3.87E-04	1.43E-03	2.91	0.00	1.99	0.00	0.92	0.00
TFEC	2.88E-05	1.46E-04	1.76	0.00	0.94	0.00	0.83	0.00
MEIS1	2.72E-05	1.45E-04	7.93	8.63	7.11	7.95	0.81	0.68
ARHGAP15	1.20E-04	4.96E-04	2.17	0.00	1.41	0.00	0.76	0.00
RUNX1	6.46E-20	6.20E-18	10.97	11.12	10.22	10.33	0.74	0.78

FLI1	1.93E-11	2.73E-10	9.27	9.81	8.53	9.10	0.74	0.70
INF2	1.44E-09	1.26E-08	7.70	7.76	7.05	7.27	0.66	0.49
ZC3H12C	1.01E-02	2.56E-02	2.57	0.00	1.92	0.00	0.65	0.00
KIT	1.57E-02	3.59E-02	5.55	6.52	4.90	5.84	0.65	0.68
PF4V1	1.49E-02	3.50E-02	3.34	2.89	2.73	2.03	0.60	0.87
BCL11A	6.80E-03	1.92E-02	1.90	0.00	1.38	0.00	0.52	0.00
GATA2	5.76E-03	1.68E-02	10.10	10.39	9.68	10.09	0.42	0.30
SMAD1	1.04E-02	2.56E-02	6.81	7.29	6.58	7.10	0.23	0.19
DYRK1A	3.92E-03	1.17E-02	8.94	8.99	8.75	8.74	0.19	0.25

Table S5. Top 40 functional enrichment terms from GREAT analysis on genes assigned to sites that were differentially bound between GATA1fl and GATA1s in G1ME cells. Enrichment analysis was performed on (i) genes assigned to 1,882 sites bound > 2-fold more by GATA1fl vs. GATA1s as well as on (ii) genes assigned to 2,612 sites bound > 2-fold more by GATA1s vs. GATA1fl. Listed are top 10 "mouse phenotype" and top 10 "GO biological processes" terms from both analyses. Based on the names of the enrichment terms, we grouped them into five categories: erythroid, megakaryocytic, myeloid, other hematopoietic, and cardiovascular and other.

Database	**Enrichment term**	**Genes bound more by GATA1fl**		**Genes bound more by GATA1s**	
		binomial FDR	**binomial fold enrichment**	**binomial FDR**	**binomial fold enrichment**
	Erythroid terms				
Mouse Phenotype	abnormal mean corpuscular volume	8.52E-20	7.35	-	-
Mouse Phenotype	abnormal erythrocyte morphology	7.81E-18	2.60	-	-
Mouse Phenotype	abnormal mean corpuscular hemoglobin	1.44E-17	8.79	-	-
Mouse Phenotype	reticulocytosis	1.66E-17	7.29	-	-
Mouse Phenotype	abnormal hemoglobin	3.66E-17	3.98	-	-
Mouse Phenotype	abnormal erythrocyte cell number	7.92E-17	3.59	-	-
Mouse Phenotype	increased red blood cell distribution width	8.25E-17	7.59	-	-
Mouse Phenotype	abnormal erythropoiesis	8.30E-17	2.41	-	-
Mouse Phenotype	hemolytic anemia	1.60E-16	14.45	-	-
GO Biological Process	erythrocyte homeostasis	9.92E-06	3.99	-	-
GO Biological Process	erythrocyte differentiation	2.13E-05	3.91	-	-
GO Biological Process	porphyrin-containing compound biosynthetic process	4.40E-05	9.60	-	-
	Megakaryocytic terms				
Mouse Phenotype	abnormal platelet physiology	-	-	8.19E-07	3.09
Mouse Phenotype	abnormal megakaryocyte morphology	-	-	2.37E-05	2.23
Mouse Phenotype	abnormal platelet activation	-	-	4.51E-05	3.37
Mouse Phenotype	abnormal megakaryocyte differentiation	-	-	6.95E-05	6.01
Mouse Phenotype	abnormal platelet aggregation	-	-	1.47E-04	3.31

Myeloid terms					
GO Biological Process	regulation of granulocyte chemotaxis	-	-	1.19E-02	5.71
GO Biological Process	positive regulation of myeloid leukocyte differentiation	-	-	1.29E-02	3.47
GO Biological Process	regulation of leukocyte migration	-	-	1.54E-02	2.59
Other hematopoietic terms					
Mouse Phenotype	abnormal lymph organ size	2.55E-17	2.17	-	-
GO Biological Process	regulation of myeloid cell differentiation	5.20E-09	6.61	-	-
GO Biological Process	negative regulation of Ras protein signal transduction	8.46E-05	6.15	-	-
Mouse Phenotype	abnormal hematopoietic system physiology	-	-	3.19E-08	2.16
Mouse Phenotype	decreased interferon-gamma secretion	-	-	2.37E-06	2.47
Mouse Phenotype	abnormal type IV hypersensitivity reaction	-	-	4.15E-06	3.29
Mouse Phenotype	increased IgG1 level	-	-	1.48E-04	2.89
GO Biological Process	positive regulation of tyrosine phosphorylation of STAT protein	-	-	1.60E-02	3.20
Cardiovascular and other terms					
GO Biological Process	fatty acid metabolic process	7.64E-05	2.45	-	-
GO Biological Process	regulation of ARF protein signal transduction	4.13E-04	4.20	-	-
GO Biological Process	homeostasis of number of cells	4.28E-04	2.43	-	-
GO Biological Process	organophosphate metabolic process	7.58E-04	2.21	-	-
GO Biological Process	progesterone receptor signaling pathway	8.77E-04	7.18	-	-
Mouse Phenotype	abnormal physiological neovascularization	-	-	1.08E-06	5.94
GO Biological Process	regulation of smooth muscle cell proliferation	-	-	1.43E-02	2.45
GO Biological Process	cardiac muscle fiber development	-	-	1.61E-02	6.73
GO Biological Process	positive regulation of osteoclast differentiation	-	-	1.63E-02	4.36
GO Biological Process	positive regulation of behavior	-	-	3.49E-02	2.06
GO Biological Process	regulation of vasoconstriction	-	-	4.51E-02	2.51
GO Biological Process	regulation of bone resorption	-	-	4.54E-02	3.45

SUPPLEMENTAL METHODS

Experimental procedures:

Cell culture

Stromal and fibroblast lines were cultured in fibroblast growth media consisting of DMEM (Mediatech), 10% fetal bovine serum, 2mM glutamine (Invitrogen), 1% penicillin/streptomycin (Gibco), 100 uM nonessential amino acids (NEAAs, Invitrogen), 0.1 mM β-mercaptoethanol (BME) and 4 ng/ml bFibroblast growth factor (bFGF, Invitrogen). Mononuclear cells (MNCs) were cultured in QBSF-60 media (Quality Biological, Inc.) supplemented with stem cell factor (SCF) 100 ng/ml, thrombopoietin (TPO) 50 ng/ml, Flt3-ligand (Flt3L) 50 ng/ml, Interleukin-3 (IL-3) 10 ng/ml, and 1% penicillin/streptomycin (Gibco). Human embryonic stem cell (hESC) media consisted of DMEM/F12 50/50 (Mediatech), 20% knockout serum replacement (Invitrogen), 2mM L-glutamine, 1% penicillin/streptomycin, 100 uM non-essential amino acids (NEAAs), 0.1 mM β-mercaptoethanol, and 10 ng/ml bFibroblast growth factor (bFGF). All iPSC lines were maintained with hESC media and on irradiated mouse embryonic fibroblasts (MEFs). Cultures were split weekly after incubation with TrypLe (Invitrogen) for 3-5 minutes and then mechanically disaggregated and plated on fresh MEFs. GATA1^{-} megakaryocyte-erythroid (G1ME) cells were maintained as described (3) in thrombopoietin (TPO)-conditioned media prepared from cells engineered to express murine TPO.

Generation and maintenance of induced pluripotent stem cells (iPSCs)

To reprogram fibroblasts and mononuclear cells, 40,000 and 200,00 cells respectively were infected with 5 microliters each of concentrated pHage2-CMV-RTTA-W and pHage-Tet-hSTEMMCA-loxP virus in the presence of 5 mcg/mL polybrene, and spinoculated at 2,250 rpm at 25°C for 1.5 hours (2, 4). One half of the media was replaced after infection. Twenty-four hours later, cells were resuspended in fresh media with 1 mcg/mL doxycycline. Cells were split onto irradiated MEFs 3-6 days after infection and colonies were picked 21-28 days after infection and expanded. After ten days, media was switched to hESC media. Doxycycline was removed after colonies appeared.

Flow cytometry

Antibodies included anti-human CD43 FITC, CD41a PE, CD42a FITC, CD235a APC or PE, CD71 APC or PE, CD18 APC, CD34 PE-Cy7, Tra-1.60 FITC, Tra-1.81 AF555 (BD Biosciences); CD31 PE-Cy7, CD45 Pacific blue, SSEA3 AF488, SSEA4 AF647 (Biolegend); VEGFR2/KDR PE (R&D Systems); CD117 (Invitrogen) and anti-mouse Ter119 APC, CD41 PE (BD Biosciences), and Gp1b PE (Emfret Analytics). Cells were stained in PBS/1%BSA at 25ºC for 20 minutes and analyzed on a FacsCanto (BD Biosciences) and with FloJo software (Tree Star, Ashlan, OR), or sorted on a FACSDiva (BD Biosciences).

Teratoma assay

For teratoma formation, 1 million feeder-depleted iPSCs were resuspended in 1:6 Matrigel (BD Biosciences) diluted in IMDM and injected intramuscularly into nonobese diabetic/severe combined immunodeficient mice. Teratomas were harvested 6-8 weeks later and paraffin sections were stained with haematoxylin and eosin. Animal experiments were performed in accordance with institutional guidelines.

Karyotyping

Karyotyping of iPSCs was performed at the Coriell Institute of Medical Research (Camden, NJ) and the Children's Hospital of Philadelphia Cytogenetics Laboratory (Philadelphia, PA).

GATA1 mutational analysis

DNA was extracted from primary patient samples and resultant iPSC clonal lines. For primary patient samples and iPSC clones with splice site mutations, *GATA1* exon 2 was amplified by PCR, fragments were cloned into the TOPO-TA vector (Invitrogen), and direct sequencing was performed using M13 standard primers, F: GTAAAACGACGGCCAG, R: CAGGAAACAGCTATGAC. For most iPSC clones, *GATA1* exon 2 was amplified by PCR, and direct sequencing was performed on the PCR product with the following primers: GATA1 exon 2, F: AAGAGGAGCAGGTGAAAGGATGTGG, R: TGACCTAGCCAAGGATCTCCATGGCAAC.

Hematopoietic differentiation by embryoid body formation

EBs were cultured in StemPro-34 (Invitrogen) media supplemented with 2 mM glutamine, 50 mcg/ml ascorbic acid, 150 mcg/ml transferrin, 0.4 mM monothioglycerol, and with bone morphogenic protein 4 (BMP4) 25 ng/ml, vascular endothelial growth factor (VEGF) 50 ng/ml (day 0-2); BMP4 25 ng/ml, VEGF 50 ng/ml, stem cell factor (SCF) 50 ng/ml, thrombopoietin (TPO) 50 ng/ml, FLT3-ligand (FLT3) 50 ng/ml, fibroblast growth factor (bFGF) 20 ng/ml (day 2-4); VEGF 50 ng/ml, SCF 50 ng/ml, TPO 50 ng/ml, FLT3 50 ng/ml, bFGF 20 ng/ml (day 4-8); SCF 50 ng/ml, TPO 50 ng/ml, interleukin-3 (IL-3) 10 ng/ml, interleukin-11 (IL-11) 5 ng/ml, erythropoietin (EPO) 2 U/ml, and insulin growth factor-1 (IGF-1) 25 ng/ml (day 8+). All cytokines except EPO (Amgen) and bFGF (Invitrogen) were purchased from R&D Systems. Cultures were maintained at 37ºC, 5% CO_2, 5% O_2, and 90% N_2.

Preparation of cells from embryoid bodies

To assay embryoid bodies (EB), suspension cells were collected from the supernatant by harvesting EB cultures and centrifuging at 600 rpm for 1 minute. To analyze total EB cultures, EBs were dissociated to single cells by a 1 hour incubation with 0.2% Collagenase B containing 20% serum followed by a 2 minute incubation with trypsin (0.05% trypsin-EDTA) at 37ºC. After enzymatic treatment, 1 ml serum was added and the EBs were disaggregated to single cells by multiple passages through a 20-gauge needle.

Hematopoietic colony-forming and liquid culture assays

CD41$^+$235$^+$ cells were seeded into H4230 methylcellulose (Stem Cell Technologies) with EPO 5 U/ml, IL-3 10 ng/ml, SCF 5 ng/ml, and granulocyte-macrophage colony stimulating factor (GMCSF) 5 ng/ml, at 2,000 - 5,000 cells/ml. Colonies were scored at 12 days. 2,000-5,000 cells/ml were seeded into Megacult-C (Stem Cell Technologies) with TPO 50 ng/ml, interleukin-6 (IL-6) 10 ng/ml, and IL-3 10 ng/ml. After 12 days, cultures were dehydrated, fixed, and stained with anti-GPIIb/IIIa antibody. For liquid culture assays, progenitor cells isolated from day 7-8 EB differentiation cultures were grown on OP9 feeder cells in serum free differentiation (SFD) medium consisting of Iscove's Mimimal Essential Media (IMDM, Life Technologies) containing 25% Ham's F12 (Cellgro) supplemented with 0.5% N2 (Life Technologies), 1% B27 without Vitamin A (Life Technologies), and 0.05% BSA diluted in PBS (Sigma). The SFD media is supplemented with 2 mM glutamine (Cellgro), 50 mg/ml ascorbic acid (Sigma), and 4 x 10^{-4} M monothioglycerol (Sigma) before use. The following cytokines were used for lineage specific cutlures: erythroid, EPO 2U/ml, SCF 100 ng/ml; megakaryocyte, SCF 100 ng/ml, TPO 50 ng/ml; and myeloid, SCF 100 ng/ml, IL-3 5 ng/ml, IL-5 5 ng/ml, and GMCSF 5 ng/ml.

Morphologic analysis

Cells were centrifuged onto a glass slide and stained with May-Grunwald-Giemsa (Sigma). Light microscopy images were obtained with a Zeiss Axioskope 2 microscope, Axiocam camera, and AxioVision 4.8 software (Carl Zeiss Microimaging).

Constructs

The human GATA1 coding sequence was cloned into the lentiviral vector HMD containing GFP to generate HMD-GATA1. The GATA1s (GATA1 lacking amino acids 1-83) mutant was amplified by PCR and inserted into HMD to generate HMD-GATA1s. The murine GATA1 coding sequence was cloned into the MSCV-based retroviral vector MIGR1 with a single HA tag (YPYDVPDYA) at the N-terminus to generate MIGR1-HA-GATA1. The GATA1s (GATA1 lacking amino acids 1-83) mutant was amplified by PCR with a single HA tag at the N-terminus and inserted into MIGR1 to generate MIGR1-HA-GATA1s.

Lentiviral transduction

The HMD lentiviral vector was used to express human wt GATA1 or truncated human GATA1s in CD41$^+$235$^+$ iPSC-derived progenitor cells. Viral particles were generated by transient transfection of 293T cells using Lipofectamine 2000 according to manufacturer's instructions, and viral supernatant collected and concentrated 100x 48 hours after transfection. For lentiviral transduction, 1.5 μl of concentrated virus was used per 1 x 10^5 cells in the presence of 2 ng/mL polybrene and 10 mM HEPES in 1 well of a 96-well plate and spun at 2250 rpm for 90 minutes at 37 °C.

Retroviral transduction

Retroviral infections of G1ME cells were carried out as described (5). The retroviral vector MIGR1 was used to express fl or mutant murine HA tagged GATA1 in G1ME cells. Viral particles were generated by transient transfection of Plat-E retrovirus packaging cells using Lipofectamine 2000 according to manufacturer's instructions, and viral supernatant collected 48 hours after transfection. For retroviral transduction, 1.2 - 1.5 mL of retroviral supernatant was mixed with 0.8 - 0.5 mL G1ME media and 2 x 10^6 cells in the presence of 8 ng/mL polybrene and 10 mM HEPES in 1 well of a 6-well plate and spun at 3200 rpm for 90 minutes at 37 °C.

Cells were incubated at 37 °C in 5% CO_2 for 3 hours and then resuspended in 5 mL fresh media. EPO 2 U/mL was added to G1ME cell transductions to support erythro-megakaryocytic differentiation.

Semi-quantitative real time polymerase chain reaction (RT-PCR) primers used:

Human RT-PCR primers (5' to 3'):

Cyclophilin
Forward GAAGAGTGCGATCAAGAACCCATGAC
Reverse GTCTCTCCTCCTTCTCCTCCTATCTTTACTT

DNMT3B
Forward TACAGACGTGTGCAGTTGTAGGCA
Reverse GTGCAGACTCCAGCCCTTGTATTT

REX1
Forward AAAGCATCTCCTCATTCATGGT
Reverse TGGGCTTTCAGGTTATTTGACT

ABCG2
Forward TCAGGAGACCACATTTCATCTAGCCC
Reverse CAGGGCACCCACTGACAAACTAAA

NANOG
Forward CCTGAAGACGTGTGAAGATGAG
Reverse GCTGATTAGGCTCCAACCATAC

For expression of lentivirus transgene OCT4-KLF4
Forward GGT GCG CCA GTA AAG CAG ACA TTA AA
Reverse CAG ACG CGA ACG TGG AGA AAG A

GATA1
Forward AGA TGA ATG GGC AGA ACA GG
Reverse ATT TCT CCG CCA CAG TGT C

BAND3
Forward TCT CTG GGA AGG TCA CAC ACC TGA
Reverse ACA CAC GGT AGG TGT GAT CCT GTT

ALAS2
Forward CCT TTG AGA CTG TCC ACT CCA
Reverse GGT GGG ACA CAT CAC ACA AC;

KLF1
Forward CAT CAG CAC ACT GAC CGC CCT G,
Reverse CAT GTC CTG CGC CTC TTC GG;

GYPA
Forward   AGG GTA CAA CTT GCC CAT CA
Reverse   ACC AGC CAT CAC CCC AAA

Murine RT-PCR primers (5’ to 3’):

Alas2
Forward   TATGTGCAGGCCATCAACTACCCA
Reverse   TTTCCATCATCTGAGGGCTGTGGT

Gp1ba
Forward   CTTGTTGCCAACGACCAAGCTGAA
Reverse   AAGCCCTTTGGTATTGTGCGAAGC

Gypa
Forward   TCACACGGCCCCTACTGAAGTGT
Reverse   TCCCTGCCATCACGCGGAAAAT

Klf1
Forward   CACGCACACGGGAGAGAAG
Reverse   CGTCAGTTCGTCTGAGCGAG

Pf4
Forward   TTCTGGGCCTGTTGTTTCTG
Reverse   GATCTCCATCGCTTTCTTCG

Thbs1
Forward   TAGCTGAGGCGGATCAGCAAATCT
Reverse   GGGAAGCCAAAGGAGTCCAAATCA

Vwf
Forward   TCATCGCTCCAGCCACATTCCATA
Reverse   AGCCACGCTCACAGTGGTTATACA

Zfpm1
Forward   CCTTGCTACCGCAGTCATCA
Reverse   ACCAGATCCCGCAGTCTTTG

ChIP qPCR primers (5’ to 3’)

Alas2 +2 kb
F’ AGGGCAGGACTTTGCCTCTAATCT
Reverse   AGATGTCCCAGTTCCTGCAGGTTT

Capn2 +13 kb
F’ TAATGGGAGTTCCCAGCATTT
Reverse   GCACAAGAGAGGATGACCTTAT

Eraf prom

F' TGCCTGCGTCTCGCTTAGT
Reverse GCTGAGCCCGCCTCATC

Ermap +1.7 kb
F' GGACAGATTCAGGAGGAGAGTA
Reverse CTTTGCACCTCTGAGCTATGAT

Fli1 prom
F' GCCCAGTTACATTCATGCAC
Reverse TGCAGACTTCAGGAATCAGG

Gp1ba prom
F' TGGTGGCTAGTAGCTGCAAAGTC
Reverse TTATCAGCTCTCTGCACAGCATTC

Gypa prom
F' GCAGTTATGCAGACCTCTAGTT
Reverse CCTCTATCCGTTGACACACATT

Hbb-b1 prom
F' CAGGGAGAAATATGCTTGTCATCA
Reverse GTGAGCAGATTGGCCCTTACC

Hbb HS3
F' CTAGGGACTGAGAGAGGCTGCTT
Reverse ATGGGACCTCTGATAGACACATCT

Itga2b prom
F' TCCTGCTCTTGAATGCTGTG
Reverse GGGAGGAAGTGGGTAAATGTC

Klf1 prom
F' TCTGCTCAAGGAGGAACAGAGCTA
Reverse GGCTCCCTTTCAGGCATTATCAGA

Lrrc39 prom
F' TTCCCTGGTGTCTGTAGGAACACA
Reverse GGGCTTCTGTGCAAAGGTTCAACT

Lyl1 prom
F' TCAGCATTGCTTCTTATCAGCC
Reverse CGCAGAGGCCAGAGGATG

Myh9 +5 kb
F' CACGATTACGGTGACCTTTCTA
Reverse CTTGACTGTGCAGAAGGAAATG

Pf4 prom
F' GCTGCTGGCCTGCACTTAAG
Reverse GCCACTGGACCCAAAGATAAAG

Src +5 kb
Forward   TTTCCTGTCCTGAAGTGGGTGGAA
Reverse   TGGATGGCTACAGCCACCTTAACT

Thbs1 -45 kb
Forward   TCACGCTGTGTTGATGAGAGCAGA
Reverse   ACTGGGTAGCAGTTCCAAGGGATT

Tubb1 +3 kb
Forward   CTGTGTTGACTTGAAGGCCTTTGG
Reverse   TGACTCCTGTGGCACATAAGGGTA

Vwf -11 kb
Forward   ATATCAGGCCTTTCCTCCAAGGGT
Reverse   GCAACTGCCTGCCATGCTATCAAT

Zfpm1 +2 kb
Forward   CTTTTCTCCTGCCCAGTCG
Reverse   TGCTGTTGCCTCGAACC

Bioinformatics analysis:

Microarray transcriptome analysis

Affymetrix HuGene 1.0 ST CEL data files were processed using RMA method implemented by the "oligo" package in R (6-8). 33,297 transcripts were collapsed to 19,392 RefSeq genes. If several transcripts mapped to one RefSeq gene, expression values were averaged to obtain one value per gene. In order to investigate differential expression on genomewide microarray data, we performed a moderated *t* test on the whole set of 19,392 genes, comparing expression of genes between T21/*GATA1s* (3 replicates) and T21/*wtGATA1* (6 replicates) iPSC-derived progenitors, using Bioconductor R "limma" (Linear Models for Microarray Data) package (9). Next, 8,519 genes that were silent (i.e. displayed $\log_2$(expression) < 5) across all 9 microarrays were filtered out from further analysis. Moderated *t* test *P* values were then corrected for multiple comparisons using BH-FDR method (10). We identified 273 differentially expressed genes (BH-FDR < 0.1), out of which 56 displayed an absolute value of $\log_2$(fold change) $\geq$ 1 (i.e. absolute value of fold change $\geq$ 2).

To identify which of the differentially expressed genes are GATA1 targets (i.e. genes that are bound by GATA1 at one or more sites within a 10 kb-gene neighborhood, i.e. 10 kb upstream of TSS + gene body + 10 kb downstream of TES), we used DNA segments occupied by GATA1 in human peripheral blood-derived erythroblasts (PBDE) and peripheral blood-derived erythroblasts from 16-19 week human fetal liver (PBDEFetal). These ChIP-seq peaks were obtained from ENCODE data (11) generated in the Snyder and Farnham labs, downloaded from the UCSC Genome Browser (12) as UCSC Accession numbers wgEncodeEH001765 and wgEncodeEH001785, and file names: wgEncodeSydhTfbsPbdeGata1UcdPk.narrowPeak.gz and wgEncodeSydhTfbsPbdefetalGata1UcdPk.narrowPeak.gz, respectively; genome assembly hg19.

Specifically, we intersected gene neighborhoods of 56 genes that were ≥ 2-fold up- or downregulated between GATA1s- and wtGATA1-expressing progenitors with GATA1 ChIP-seq peaks from PBDE and PBDEFetal cell lines. Among the 19,392 RefSeq genes represented on Affymetrix HuGene 1.0 ST microarray, 8,839 were occupied by GATA1 in PBDE and/or PBDEFetal cell lines. Therefore, an expected fraction of GATA1s targets in a randomly sampled set of genes is ~45%. We found that 19 out of 22 genes (~86%) downregulated in *GATA1s* as compared to *wtGATA1* progenitors were bound by GATA1 (Table S2), and thus were likely GATA1 targets. This corresponds to ~2-fold enrichment over what is expected by chance (binomial test *P* value = 10^{-4}). Conversely, although 20 out of 34 genes (59%) upregulated in T21/*GATA1s* cells were bound by GATA1 in human PBDE and/or PBDEFetal cells (Table S2), this is not significantly different from random expectation (1.3-fold enrichment; binomial test *P* value = 0.084).

To further characterize differentially expressed genes we used GSEA (13, 14) in which we utilized microarray expression data for several human hematopoietic cell populations from Novershtern *et al.* (1). Specifically, to define erythroid, myeloid, and megakaryocytic transcriptome signature we used microarray data from the following cell populations: (i) erythroid signature: Erythroid CD34-CD71lowGlyA+ cells (7 replicates), and Erythroid CD34-CD71-GlyA+ cells (6 replicates); (ii) myeloid signature: basophils (6 replicates), eosinophils (5 replicates), and neutrophils (4 replicates); and (iii) megakaryocytic signature: CFU-megakaryocytes (5 replicates), and megakaryocytes (7 replicates). We performed two GSEAs: (i) GSEA on all 273 differentially expressed genes (154 genes up- and 119 genes downregulated in GATA1s- as compared to wtGATA1-expressing progenitors, BH-FDR < 0.1) (Figure 4B); and (ii) GSEA on 56 differentially expressed genes (34 up- and 22 downregulated genes) that not only pass the BH-FDR threshold of < 0.1, but also display an absolute value of $\log_2$(fold change) ≥ 1 (Figure S4). To assess the significance of enrichment scores we performed 1000 "phenotype" permutations. Processing of human euploid microarray samples and analysis done to generate Figure S5 were performed analogously to what is described above for trisomy 21 samples. Heat maps in Figures 4, S5, and S7 were prepared using HeatMapImage module at http://genepattern.broadinstitute.org/ (15).

Mouse G1ME transcriptome analysis using Affymetrix Mouse Genome 430 2.0 microarrays, comparing mean expression of genes in G1ME/*GATA1s* vs. G1ME/*GATA1fl* (3 replicates each), was done as described above for human iPSC-derived progenitor samples with few differences. RMA processing was done using "affy" package in R (16). After RMA processing 45,101 probesets were mapped to 21,246 genes. After the moderated *t* test was performed and before the BH-FDR multiple test correction was applied, 4,916 genes that were silent (i.e. displayed $\log_2$(expression) < 4) across all 6 microarrays were removed from further analysis. We identified 61 genes that were significantly downregulated and 75 genes that were significantly upregulated (FDR < 0.1 & absolute value of fold change ≥ 2) in G1ME/*GATA1s* vs. G1ME/*GATA1fl*.

Single cell gene expression analysis

Expression levels for 94 selected genes (Table S3), including 91 key hematopoietic genes and 3 housekeeping genes, in single cells purified by flow cytometry from iPSC differentiation cultures, were measured by quantitative RT-PCR. Cycle threshold (Ct) numbers were downloaded from the Fluidigm BioMark software and used to calculate relative $\log_2$(expression level) of analyzed genes in single cells using the following formula from the Fluidigm "Application Guidance: Single-Cell Data Analysis" manual as per manufacturer instructions:

$$log_2(G_i) = LOD - Ct_i$$

where G_i is the relative expression of gene *i*, *LOD = 24* is the selected limit of detection, and Ct_i is the cycle threshold number of gene *i*. If value is negative, $log_2(G_i)$ = *0.*

Cells with low expression (i.e. > 3 standard deviations away from the median expression across all single cells analyzed) of two housekeeping genes, ACTB and GAPDH, were excluded from further analyses, resulting in a total of 755 single cells. These cells included CD41$^+$235$^+$ progenitors from *GATA1s* (n = 274) or *wtGATA1* (n = 311) iPSCs. As controls, lineage-committed erythroid (CD41$^-$235$^+$, n = 57), megakaryocytic (CD41$^+$42$^+$, n = 61), and myeloid (CD45$^+$18$^+$, n = 52) cells were examined (Figure 5A). Data was normalized using additive correction on the logarithmic scale so that all cells have the same median $\log_2$(expression value) of detected genes (i.e. genes whose *$log_2(G_i)$* is > 0) equal to the average median expression across all cells.

PCA was performed using princomp(x) function in MATLAB on single cell gene expression data for erythroid, megakaryocytic, and myeloid reference cells. This resulted in an identification of the first two principal components – PC1 and PC2 – which accounted for 65% of the variance in the data and resulted in clear clustering of the three reference cell types (Figure 5B, left). To project the expression patterns of *wtGATA1* and *GATA1s* progenitor cells onto PC1 and PC2 plane identified for committed cells, we first shifted the progenitor expression data using the shift that was applied to committed cells during data centering for PCA. We then used PC1 and PC2 loadings (coefficients) obtained from PCA on committed cells to calculate PC1 and PC2 scores, i.e. projection of expression patterns of each progenitor cell onto PC1 and PC2 (Figure 5B). To investigate whether GATA1s- and wtGATA1-expressing progenitors differ significantly along the PC1 direction we performed a Mann Whitney *U* test comparing PC1 score distributions of these two populations of cells.

We used PC1 and PC2 scores shown in Figure 5B, as predictor variables in Linear Discriminant Analysis (LDA) to classify *wtGATA1* and *GATA1s* progenitors into erythroid, myeloid, or megakaryocytic lineage. Specifically, we used projections of committed cells onto PC1 and PC2 to train an LDA model that discriminates the three lineages (lda(x) function from MASS package in R (17)). The trained model was 98% correct in classifying the committed cells into their respective lineages. We then used this model to assign, to each progenitor cell expressing wtGATA1 or GATA1s, probabilities of belonging to an erythroid, myeloid, or megakaryocytic lineage. Biologically, these probabilities can be used to approximate the likelihood with which a particular progenitor cell will differentiate towards a given lineage. Next, we assigned each progenitor cell into one of four categories: erythroid, megakaryocytic, or myeloid using a probability threshold of > 0.90, or unclassified if all three probabilities assigned to a cell were < 0.90.

For each of the 94 genes interrogated, we also analyzed distributions of expression levels among single cells (see violin plots in Figures 5D and S6). Specifically, to investigate whether these distributions differed significantly between populations of single iPSC-derived progenitors expressing GATA1s or wtGATA1, the non-parametric Mann-Whitney *U* test was performed, followed by a multiple comparison correction using BH-FDR method (10).

Violin plots in Figures 5D and S6 were prepared using “vioplot” package in R (18). Hierarchical clustering heat maps in Figure 6B were prepared using heatmap.2(x) from “gplots” package in R (19). Hierarchical clustering was performed using complete linkage method and Euclidean measure of distance.

Genomewide differential binding analysis on ChIP-seq data.

We performed ChIP-seq analysis on GATA1fl (2 replicates) and GATA1s (2 replicates) in G1ME cells at 42 hours post-transduction. For GATA1fl samples, we called 24,579 peaks in replicate 1 and 14,328 peaks in replicate 2, with 9,205 peaks present in both GATA1fl replicates. For GATA1s samples, we called 26,024 peaks in replicate 1, and 28,420 peaks in replicate 2, with 13,657 peaks present in both GATA1s replicates. Differential binding analysis was performed using “DiffBind” package in R (20, 21) on GATA1fl and GATA1s binding sites that were called as peaks in both respective replicates (9,205 peaks for GATA1fl and 13,657 peaks for GATA1s). In total 16,231 binding sites, representing a union of GATA1fl and GATA1s peaks (after merging of peaks that overlap between GATA1fl and GATA1s), were included in differential binding analysis. To remove background noise, control read counts from matching input samples were subtracted from respective ChIP-seq samples before analysis. Read counts obtained for each of 4 replicates at 16,231 binding sites were normalized using “effective library size”, i.e. number of reads within peaks. Differential binding analysis was performed using edgeR method implemented by “DiffBind”. Binding sites were called as differentially bound using FDR threshold of < 0.1. Differentially bound sites with > 2-fold change in binding signal were assigned to predicted target genes using GREAT (22). This was achieved using “single nearest gene within 1Mb” option for associating genomics regions with genes.

SUPPLEMENTAL REFERENCES

1. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell.* 2011;144(2):296-309.
2. Chou ST, Byrska-Bishop M, Tober JM, Yao Y, Vandorn D, Opalinska JB, Mills JA, Choi JK, Speck NA, Gadue P, 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):17573-8.
3. Stachura DL, Chou ST, and Weiss MJ. Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. *Blood.* 2006;107(1):87-97.
4. Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, and Mostoslavsky G. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells.* 2009;27(3):543-9.
5. Chou ST, Khandros E, Bailey LC, Nichols KE, Vakoc CR, Yao Y, Huang Z, Crispino JD, Hardison RC, Blobel GA, et al. Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate. *Blood.* 2009;114(5):983-94.
6. Carvalho BS, and Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics.* 2010;26(19):2363-7.
7. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80.

8. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, and Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 2003;31(4):e15.

9. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3(Article3.

10. Benjamini Y, and Hochberg Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological.* 1995;57(289-300).

11. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.

12. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* 2013;41(Database issue):D56-63.

13. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34(3):267-73.

14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, 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):15545-50.

15. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, and Mesirov JP. GenePattern 2.0. *Nat Genet.* 2006;38(5):500-1.

16. Gautier L, Cope L, Bolstad BM, and Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004;20(3):307-15.

17. Venables WN, and Ripley BD. *Modern Applied Statistics.* Spinger; 2002.

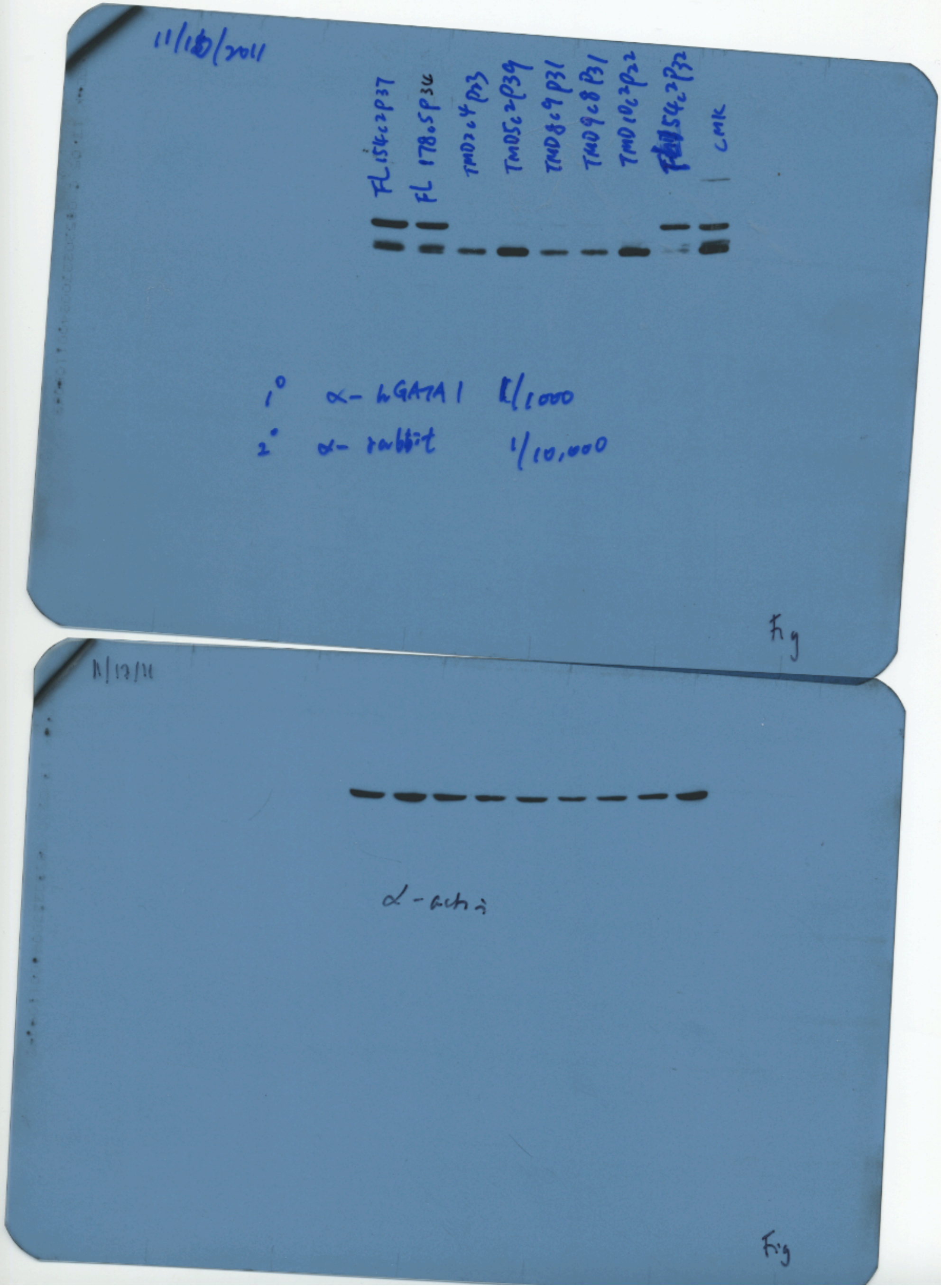
18. Adler D. Violin Plot. R package version 2.0. http:/wsopuppenkiste.wiso.uni-goettingen.de/~dadler; 2005.

19. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, et al.; 2014.

20. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 2012;481(7381):389-93.

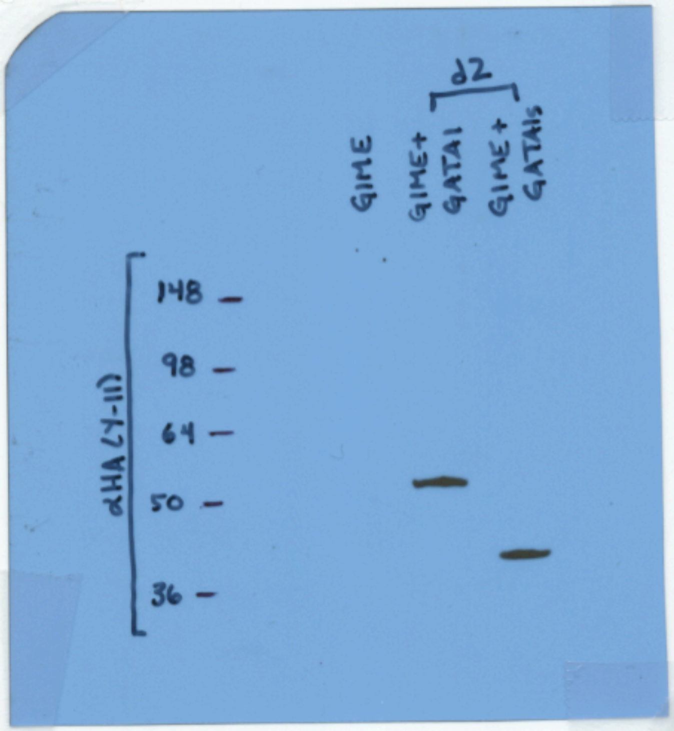
21. Stark R, and Brown G. Diffbind: differential binding analysis of ChIP-seq peak data. Bioconductor 2011; http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

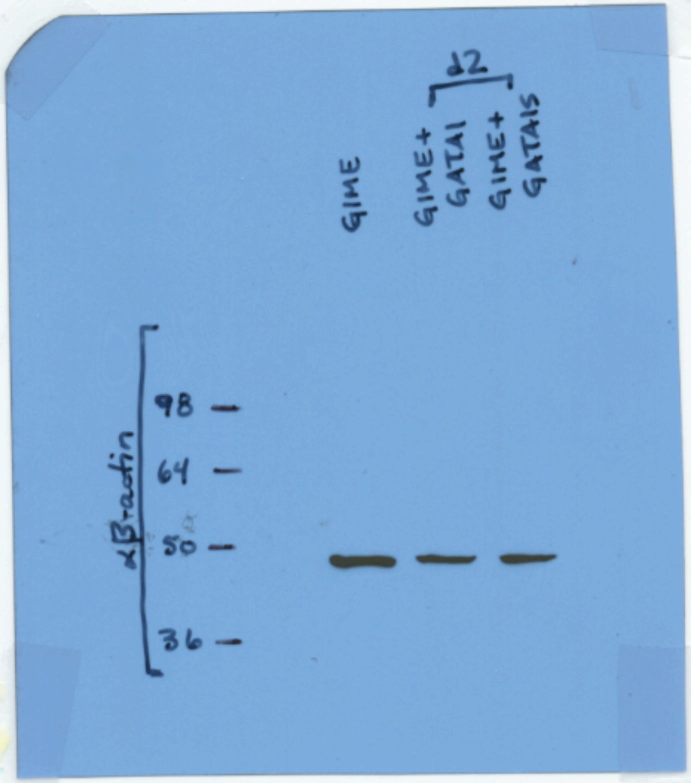
22. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, and Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28(5):495-501.



SUPPLEMENTAL
UNCUT GELS

Lane
1 = WT
2 = T21
3 = 2.4
4 = 5.2
5 = 8.9
6 = 9.8
7 = 10.2
8 not shown
9 not shown


d2
GIME
GIME+ GATA1
GIME+ GATA1s
αHA (Y-11)
148
98
64
50
36


d2
GIME
GIME+ GATA1
GIME+ GATA1s
αβ-actin
98
64
50
36