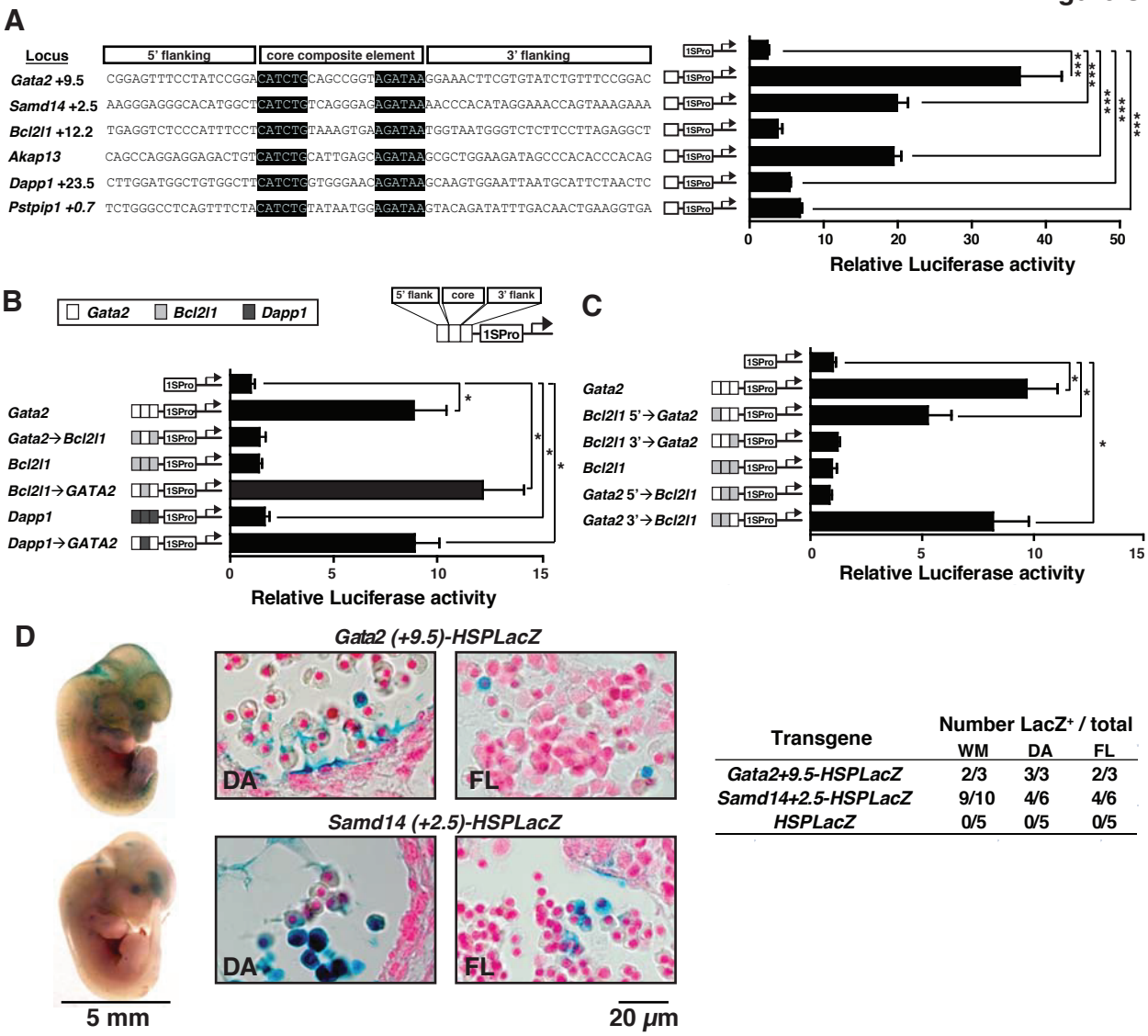
Figure S1



Transgene	Number LacZ⁺ / total		
	WM	DA	FL
Gata2+9.5-HSPLacZ	2/3	3/3	2/3
Samd14+2.5-HSPLacZ	9/10	4/6	4/6
HSPLacZ	0/5	0/5	0/5

Figure S2

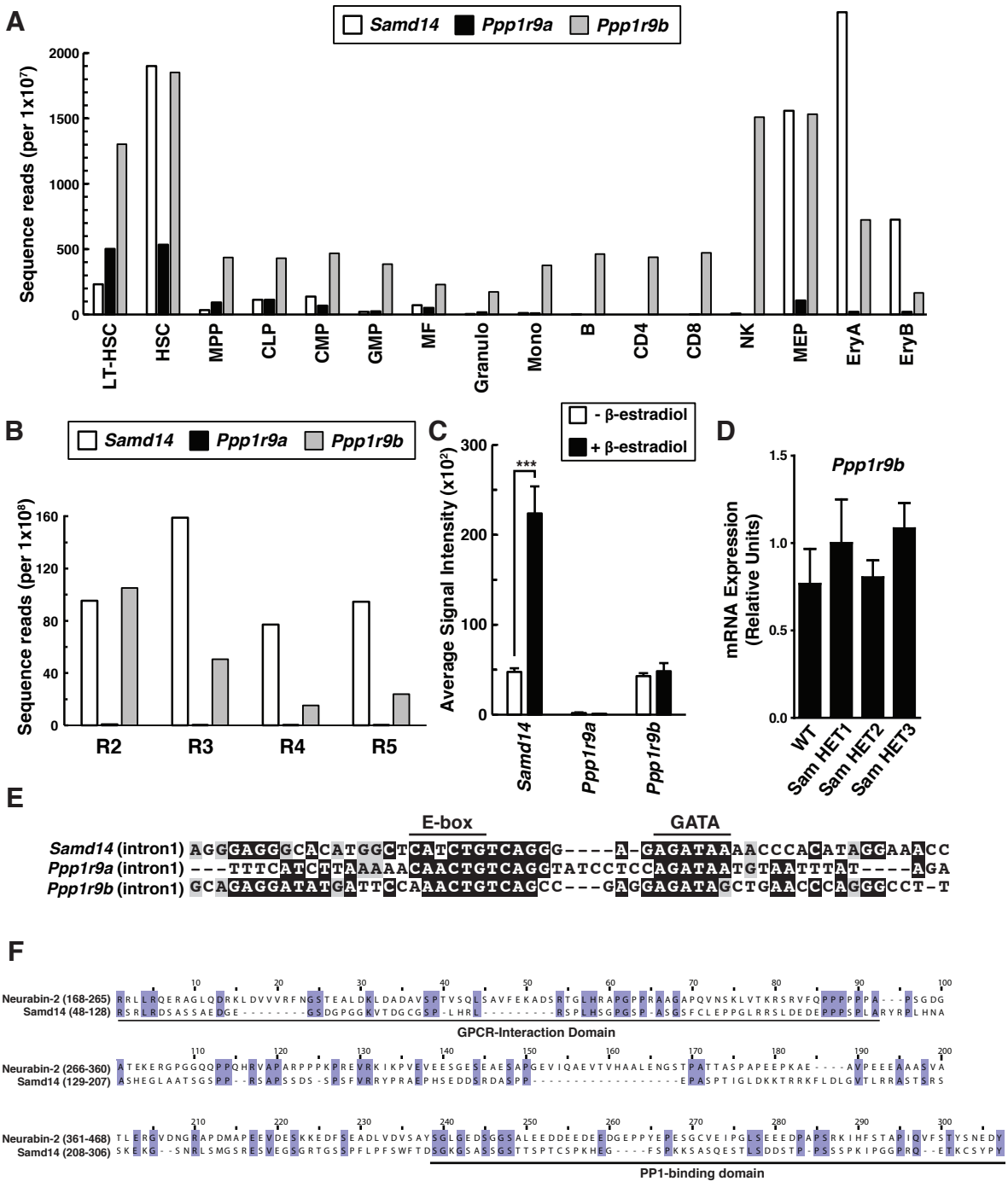
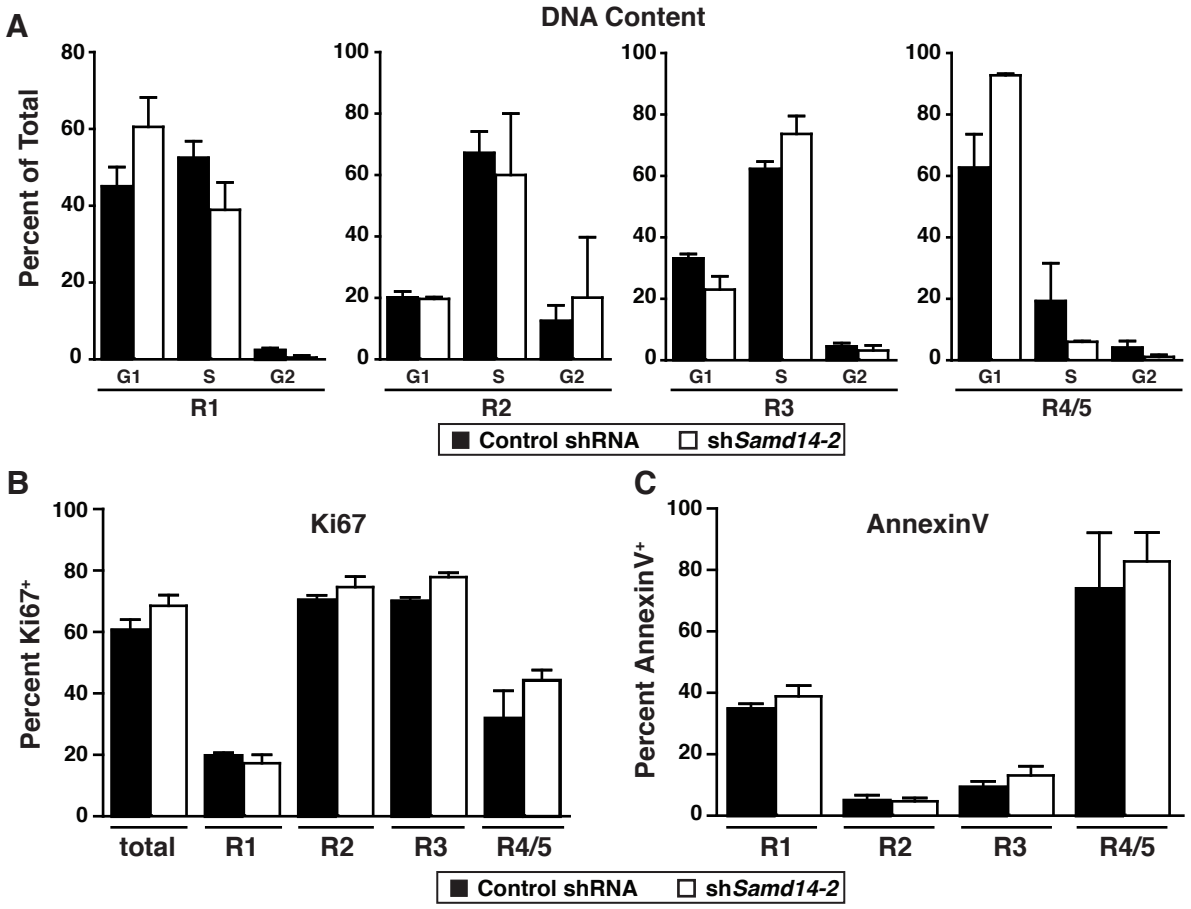

A
Samd14
Ppp1r9a
Ppp1r9b
Sequence reads (per 1x10^7)
2000
1500
1000
500
0
LT-HSC
HSC
MPP
CLP
CMP
GMP
MF
Granulo
Mono
B
CD4
CD8
NK
MEP
EryA
EryB
B
Samd14
Ppp1r9a
Ppp1r9b
Sequence reads (per 1x10^8)
160
120
80
40
0
R2
R3
R4
R5
C
- β-estradiol
+ β-estradiol
Average Signal Intensity (x10^2)
300
200
100
0
***
Samd14
Ppp1r9a
Ppp1r9b
D
Ppp1r9b
mRNA Expression (Relative Units)
1.5
1.0
0.5
0.0
WT
Sam HET1
Sam HET2
Sam HET3
E
E-box
GATA
Samd14 (intron1) AGGGAGGGCACATGGCTCATCTGTCAGGG----A-GAGATAAAACCCACATAGCAAACC
Ppp1r9a (intron1) ---TTTCATCTTAAAAACAACTGTCAGGTATCCTCCAGATAATGTAATTTAT----AGA
Ppp1r9b (intron1) GCAGAGGATATGATTCCAAACTGTCAGCC---GAGGAGATAGCTGAACCCAGGCCCT-T
F
Neurabin-2 (168-265) RRLLRQERAGLQDRKLDVVVRFNGSTEALDKLDADAVSPTVSQLSAVFEKADSRTGLHRAPGPPRAAGAPQVNSKLVTKRSRVFQPPPPPPA---PSGDG
Samd14 (48-128) RSRLRDSASSAEDGE--------GSDGPGGKVTDGCGSP-LHRL----------RSPLHSGPGSP-ASGSFCLEPPGLRRSLDEDEPPPSPLARYRPLHNA
GPCR-Interaction Domain
Neurabin-2 (266-360) ATEKERGPGGQQPPQHRVAPARPPPKPREVRKIKPVEVEESGESEAESAPGEVIQAEVTVHAALENGSTPATTASPAPEEPKAE----AVPEEEAAASVA
Samd14 (129-207) ASHEGLAATSGSPP--RSAPSSDS-SPSFVRRYPRAEPHSEDDSRDASPP----------EPASPTIGLDKKTRRKFLDLGVTLRRASTSRS
Neurabin-2 (361-468) TLERGVDNGRAPDMAPEEVDESKKEDFSEADLVDVSAYSGLGEDSGGSALEEDDEEDEEDGEPPYEPESGCVEIPGLSEEEDPAPSRKIHFSTAPIQVFSTYSNEDY
Samd14 (208-306) SKELKG--SNRLSMGSRESVEGSGRTGSSPFLPFSWFTDSGKGSASSGSTTSPTCSPKHEG----FSPKKSASQESTLSDDSTP-PSSSPKIPGGPRQ--ETKCSYPY
PP1-binding domain

Figure S3


A
DNA Content
Percent of Total
G1
S
G2
R1
R2
R3
R4/5
Control shRNA
shSamd14-2
B
Ki67
Percent Ki67+
total
R1
R2
R3
R4/5
C
AnnexinV
Percent AnnexinV+

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Enhancer Function of +9.5-like Elements *In Vivo*, Related to Figure 2

(A) Enhancer activity assay with +9.5-like sites using 65 bp encompassing the composite element cloned upstream of luciferase and transfected into G1E cells. (B) Chimeric enhancers with swapped spacer regions from the active +9.5 into the inactive *Bcl2l1* (and vice versa). (C) Chimeras containing 18 bp 3' downstream of the +9.5 element replacing the inactive *Bcl2l1* element. (D) Enhancer activity of a 65 bp *Samd14* +2.5-*LacZ* construct vs. 65 bp of the +9.5 site in E12.5 transgenic mouse embryos. (Statistical significance represented by mean +/- SEM.; *p<0.05. **p<0.01, ***p<0.001.)

Figure S2. Relationship of Samd14 to Neurabin-1 and Neurabin-2, Related to Figure 4

(A) *Samd14*, *Ppp1r9a* (neurabin-1), and *Ppp1r9b* (neurabin-2) expression in distinct hematopoietic cell subtypes. (B) *Samd14*, *Ppp1r9a,* and *Ppp1r9b* expression in fetal liver erythroid precursors. (C) *Samd14*, *Ppp1r9a,* and *Ppp1r9b* expression in G1E-ER-GATA1 cells pre- and post-b-estradiol treatment. (D) *Ppp1r9b* expression in wild type and *Samd14* +2.5 heterozygous mutant clonal lines (HET1-3 represent 3 distinct clonal lines). (E) Alignment of GATA motifs in intron 1 of *Samd14*, *Ppp1r9a* and *Ppp1r9b*. (F) Samd14 and neurabin-2 protein sequence comparison and pairwise alignment.

Figure S3. Cell cycle Analysis of Control and Samd14 Knockdown Fetal Liver Cells, Related to Figure 4

(A) FACS-purified fetal liver R1-R4/5 cells quantitated for DNA content. (B) Ki67 staining in FACS-purified fetal liver R1-R4/5 cells. (C) Quantitation of AnnexinV^{+} cells with live fetal liver cells.

SUPPLEMENTAL TABLES

Table S1. Attributes of +9.5 and +9.5-like Composite *Cis*-Elements, Related to Figure 1

Table S2. GATA-2 and Scl/TAL1 Occupancy of +9.5 and +9.5-like Elements Based on ChIP-seq Analysis, Related to Figure 3

Table S3. Influence of *Samd14* Knockdown on G1E Cell Transcriptome, Based on RNA-Seq Analysis, Related to Figure 4

EXTENDED EXPERIMENTAL PROCEDURES

Bioinformatics

We ranked 797 loci (mm9), matching the +9.5 sequence CATCTG-N[8]-AGATAA, based on their similarity to the +9.5 using multiple sources of genomic data (76 mouse ENCODE histone modifications and 125 transcription factor ChIP-seq datasets (Dore et al., 2012; Mouse et al., 2012; Tijssen et al., 2011; Wu et al., 2011)). We generated binary feature vectors for each locus based on individual data sources by overlapping the loci coordinates with the peak coordinates from the ChIP-seq datasets. We evaluated the binary Jaccard distances of each locus to the +9.5 based on each data resource and aggregated the distances to generate a dissimilarity metric. The 797 +9.5-like loci were ranked with respect to this metric and annotated based on proximity to mm9 Refseq genes (Table S1).

To identify conserved elements between mouse (mm9) and human (hg19), we annotated sequence, gene location and information for each of 102,427 hg19 candidate *cis*-elements. For lift-over conservation analysis, we used UCSC's Lift Genome Annotation tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver), to identify human elements that overlapped with mouse elements when lifted to the mouse genome, either directly or extended by 10 kb in either direction from the lifted-over hg19 element. For location-based conservation, composite elements were deemed conserved if composite elements were located in the same region of a locus between hg19 and mm9 genomes.

To assess GATA-2 and Scl/TAL1 occupancy of the 797 mouse mm9 elements, and conservation of GATA-2 and Scl/TAL1 occupancy in human hg19, we utilized ChIP-Seq peaks obtained from wide-ranging and comprehensive ChIP-seq analyses in

hematopoietic cells. Mouse GATA-2 occupancy - (Dore et al., 2012; Li et al., 2011; Trompouki et al., 2011; Wilson et al., 2010); mouse Scl/TAL-1 occupancy - (Soler et al., 2010; Wilson et al., 2010; Wu et al., 2011); human GATA-2 occupancy - (Beck et al., 2013; Fujiwara et al., 2009); human Scl/TAL1 - (Beck et al., 2013), Snyder Lab ENCODE data. Conserved occupancy was based on sequence annotation and gene location for each mm9 element and hg19 candidate *cis*-element overlapping with GATA-2 and Scl/TAL1 peaks in corresponding human and mouse datasets.

RT-PCR

Total RNA was purified from 2-3 x 10^6 cells in 1 ml TRIzol (Life Technologies). For primary transcripts, RNA was subjected to on-column DNase treatment and purification (Qiagen RNeasy). To synthesize cDNA, 2 µg RNA was incubated with 250 ng of random and oligo(dT) primers preheated at 68°C for 10 min. RNA was incubated with reverse transcriptase (Life Technologies) with 10 mM dithiothreitol (DTT), RNAsin (Promega), and 0.5 mM deoxynucleoside triphosphates at 42°C for 1 h, and then heat inactivated at 98°C for 5 min. Real-time PCR reactions contain 2 µl of cDNA, 10 µl SYBR Green (Applied Biosystems), and 100 to 500 nM of the appropriate primers, and each biological replicate sample was analyzed in duplicate on a ViiA 7 real-time PCR instrument (Applied Biosystems). Relative expression levels were determined using a standard curve of serial dilutions from appropriate control cDNA (for mRNA analysis) or sonicated genomic DNA (for allele-specific primary transcript analysis) for each primer pair. For primary transcripts, samples lacking RT, as well as samples lacking the mutant allele sequence, were negative controls.

Cell culture and generation of TALEN and CRISPR/Cas9-deleted cells

G1E cells were maintained in Iscove's modified Dulbecco's medium (GIBCO) supplemented with 15% FBS (Gemini), 1% penicillin/streptomycin (Gemini), 2 U/ml erythropoietin, 120 nM monothioglycerol (Sigma), and 0.6% conditioned medium from an SCF-producing CHO cell line. HPC-7 cells were grown in IMDM supplemented with 10% FBS, 10% CHO-KLSF conditioned medium, 1X pen/strep, and $7.5x10^{-5}$M monothioglycerol. Cell cultures were maintained in a 37°C incubator with 5% CO_2. TALENs were generated as described (Kim et al., 2013), and sgRNAs were generated by hemi-nested PCR-amplified construction of a U6 promoter-driven sgRNA, which was blunt-end cloned into SmaI-cut pBluescript (Addgene). 10 µg sgRNA-containing plasmids were co-nucleofected into 3 x 10^6 G1E cells with Cas9-expressing plasmid using Amaxa Kit R (Lonza). 72 h post-transfection, cells were cloned at limiting dilution in a 48 well plate. Cells were screened after 1 week to detect mutations using the T7 endonuclease test (Cho et al., 2013; Kim et al., 2009). DNA was amplified by PCR, denatured, reannealed to facilitate heteroduplex formation, incubated with T7 Endonuclease I (New England Biosystems) for 15 min. Clones containing target mutations were sequence-validated and subcloned to ensure clonality. Validation of allele-specific primers was conducted using template or mutant cell cDNA.

Reporter plasmids and luciferase assay

The sequence for *Gata2*-1S promoter was amplified from murine DNA, and cloned into PGL3-basic (Promega) luciferase reporter plasmid. Reporter plasmids were generated by synthesizing short oligos and directionally cloning into PGL3-1S-luciferase using

NheI and XhoI restriction enzymes, confirmed by sequencing. *LacZ* reporter constructs were generated by cloning short oligos into the pWHERE vector, containing 700 bp of the HSP70 promoter and flanking H19 insulator elements to minimize positional effects of transgene integration.

G1E cells were transfected with 5 µg of reporter DNA and 0.5 µg of Renilla luciferase diluted in 500 µl Opti-MEM reduced-serum media. 5 µl DMRIE-C was diluted in 500 µl Opti-MEM, then mixed with the DNA solution and incubated at 25°C for 30 minutes. G1E cells were harvested by centrifugation and resuspended in OptiMEM at 5 x 10^6 cells/ml. One million cells were added per condition and incubated for 4.5 hours at 37°C. 2 ml growth media containing IMDM with 15% FBS was added to the cells, and culture plates were incubated at 37°C for 48 hours. Cells were harvested by centrifugation and analyzed for luciferase activity using the Dual-Glo Luciferase (Promega) on a Lumax luminometer. Six biological replicates were analyzed for each condition. Statistics were calculated using unpaired Student's t-test.

To generate *LacZ* reporter mice, reporters containing 65 bp of *Gata2* +9.5 or *Samd14* +2.5 composite elements were linearized using PacI and gel purified. Microinjection of FVB strain fertilized eggs was performed by the U. Wisconsin-Waisman Center. Embryos were genotyped by construct-specific PCR using purified DNA obtained from the yolk sac. Control and transgene-positive embryos were fixed with 2% paraformaldehyde, 0.2% glutaraldehyde, and 0.02% IGEPAL for 2.5 hours at 4°C, then washed 3 times in PBS. For whole mount imaging, embryos were incubated at 37°C for 16 h in a solution containing 0.5 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal [Sigma]) (Wozniak et al., 2007). Embryos were post-fixed with

4% formaldehyde and 0.2% glutaraldehyde for 24 h, dehydrated, and oriented head-down in paraffin blocks. 8 µm sections were generated, and slides were counterstained with 0.1% nuclear fast red solution (EMS).

Western Blotting

Proteins were boiled in SDS buffer (25 mM Tris, pH 6.8, 2% β-mercaptoethanol, 3% SDS, .005% bromophenol blue, 5% glycerol) for 10 min, resolved by SDS-PAGE, and detected with ECL Plus (GE Healthcare). Antibodies: anti-HA (Covance HA11), anti-c-Kit (Cell Signaling Technology), anti-CD71 (Santa Cruz), anti-α-tubulin (Calbiochem). Secondary antibodies goat-anti-mouse-IgG-HRP, goat-anti-rabbit-IgG-HRP, goat-anti-rat-IgG-HRP (Santa Cruz Biotechnologies). For cell surface protein biotinylation, $Ter119^+$ cells were depleted by negative-selection bead sorting and incubated in 2 mM Sulfo-NHS-Biotin (Pierce), then in 100 mM glycine, and lysed in RIPA buffer. Biotinylated protein lysates were pulled down with EZview™ Red Streptavidin Affinity Gel (Sigma), boiled in SDS buffer, and analyzed by SDS-PAGE.

Flow cytometry

For lineage-differentiation analysis of primary fetal liver cells, $2x10^6$ cells were harvested, washed in PBS and stained using 0.8 µg each of CD71/PE (eBiosciences) and Ter119/APC (eBiosciences) in 100 µl of 2% FBS, with EDTA and glucose for 30 min at 4°C. Cells were washed again, resuspended in 500 µl PBS and DAPI was added as a live/dead cell stain. For c-Kit surface staining, cells were incubated for 30 min with Biotin-c-kit (BD Bioscience 1:200), washed and stained with PECy7-step (eBioscience

1:400), CD71/PE, and Ter119/APC antibodies for 30 min at room temp. Cell cycle analysis involved overnight fixation in 90% ethanol and staining with 40 μg/ml propidium iodide at -20°C. Apoptosis was assayed by staining with Alexa Fluor 350-conjugated Annexin V (Invitrogen). Proliferation was assayed by staining fixed cells with FITC-Ki67 (BD Biosciences). Samples were analyzed in triplicate on an LSR-II instrument (BD Biosciences). To detect intracellular phosphorylated AKT, fetal liver cells were sorted using magnetic beads for $Ter119^-$ cells (STEMCELL Technologies), serum-starved in IMDM with 1% BSA for 1 h at 37°C, stimulated with 10 ng/ml of SCF for 10 min at 37°C, and immediately fixed in 2% paraformaldehyde for 10 min at 37°C. Cells were permeabilized overnight in 95% methanol, incubated in HBSS with 4% FBS for 1 h at 4°C, and stained with phospho-AKT (1:200), biotin-c-Kit, and CD71/PE for 30 min at room temperature, and measured on an LSR II. All data was generated in triplicate and normalized to unstimulated cells.

Colony Formation

E14.5 fetal liver cells were lineage-depleted and infected with control shRNA- or shSamd14-containing virus and incubated overnight at 37°C with 5% CO_2. Cells (0.3 ml at 2 x 10^5/ml) were mixed with 3 ml of MethoCult™ GF M3434 (STEMCELL Technologies) and 1.1 ml was plated in replicate 35 mm dishes, and biological triplicates were analyzed for each condition. Colonies were enumerated and imaged for GFP+ colonies 8 days after culturing.

RNA-sequencing

Control-infected and shSamd14-infected fetal liver cells were FACS-sorted on a FACSAria III flow cytometer (BD Biosciences), and $2x10^5$ R1 cells ($CD71^{low}$, $Ter119^-$) were isolated. RNA was purified in Trizol (Life Technologies). Samples were analyzed on an Illumina HiSeq 2500. Transcript quantification and differential expression were conducted using RSEM and DESeq2, respectively. RSEM v1.2.15 was provided with a reference transcript set consisting of all protein coding and large intergenic non-coding transcripts from the Ensembl release 75 annotation of the NCBI Build 38 mouse genome assembly. Default parameters and Bowtie v1.0.1 were used for transcript quantification with RSEM. Transcript read counts were summed to produce counts at the level of gene symbols. These counts were then given as input to DESeq2 v1.2.10 for differential expression analysis. DESeq2 was run with default parameters. Gene symbols with Benjamini–Hochberg FDR values less than 0.05 were deemed to be differentially expressed between mutant and wild-type The Database for Annotation, Visualization and Integrated Discovery (DAVID) (david.abcc.ncifcrf.gov/) and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (http://string-db.org/) were used to analyze functional terms and gene associations.

Retroviral infections

Two distinct shRNAs targeting exon 8 and exon 10 of *Samd14* and *Lbd1* transcripts were cloned into MSCV-PIG (IRES-GFP) (from Mitchell Weiss) and packaged into retrovirus by cotransfection with pCL-Eco packaging vector in 293T cells. Retroviral infections were conducted following generation of Lin^- fetal liver cells. Cells (5×10^5) were added to 100 µl viral supernatant, polybrene (8 µg/ml) and HEPES buffer (10

µl/ml), and spinoculated at 1200 *x* g for 90 min at 30°C. Samples for flow cytometry, gene expression and Western blotting were analysed 72 h post-cell isolation.

Statistics

For quantitation of mRNA, primary transcript, cell number, cell colonies, and fluorescence intensities, the results are presented as mean +/- SEM. All analyses were conducted with GraphPad Prism. p-values representing confidence in mean expression value differences were calculated using unpaired Student's t-test with a cutoff of $p<0.05$ deemed significant. The q-value cutoff of <0.05 was used to define significant changes for RNA-seq data, defined as the maximum false discovery rate at which an observed score is deemed significant.

Beck, D., Thoms, J.A., Perera, D., Schutte, J., Unnikrishnan, A., Knezevic, K., Kinston, S.J., Wilson, N.K., O'Brien, T.A., Gottgens, B.*, et al.* (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood *122*, e12-22.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature biotechnology *31*, 230-232.

Dore, L.C., Chlon, T.M., Brown, C.D., White, K.P., and Crispino, J.D. (2012). Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. Blood *119*, 3724-3733.

Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol Cell *36*, 667-681.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res *19*, 1279-1288.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E.*, et al.* (2013). A library of TAL effector nucleases spanning the human genome. Nature biotechnology *31*, 251-258.

Li, L., Jothi, R., Cui, K., Lee, J.Y., Cohen, T., Gorivodsky, M., Tzchori, I., Zhao, Y., Hayes, S.M., Bresnick, E.H.*, et al.* (2011). Nuclear adaptor Ldb1 regulates a

transcriptional program essential for the maintenance of hematopoietic stem cells. Nat Immunol *12*, 129-136.

Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R.*, et al.* (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol *13*, 418.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van Ijcken, W.*, et al.* (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev *24*, 277-289.

Tijssen, M.R., Cvejic, A., Joshi, A., Hannah, R.L., Ferreira, R., Forrai, A., Bellissimo, D.C., Oram, S.H., Smethurst, P.A., Wilson, N.K.*, et al.* (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev Cell *20*, 597-609.

Trompouki, E., Bowman, T.V., Lawton, L.N., Fan, Z.P., Wu, D.C., DiBiase, A., Martin, C.S., Cech, J.N., Sessa, A.K., Leblanc, J.L.*, et al.* (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell *147*, 577-589.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E.*, et al.* (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell *7*, 532-544.

Wozniak, R.J., Boyer, M.E., Grass, J.A., Lee, Y., and Bresnick, E.H. (2007). Context-dependent GATA factor function: combinatorial requirements for transcriptional control in hematopoietic and endothelial cells. J Biol Chem *282*, 14665-14674.

Wu, W., Cheng, Y., Keller, C.A., Ernst, J., Kumar, S.A., Mishra, T., Morrissey, C., Dorman, C.M., Chen, K.B., Drautz, D.*, et al.* (2011). Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. Genome Res *21*, 1659-1671.