Extended Experimental Procedures for

Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration

Eirini Trompouki[1], Teresa V. Bowman[1*], Lee N. Lawton[2], Zi Peng Fan[2,6], Dai-Chen Wu[3,4], Anthony DiBiase[1], Corey S. Martin[1], Jennifer N. Cech[1], Anna K. Sessa[1], Jocelyn L. Leblanc[1], Pulin Li[1], Ellen Durand[1], Christian Mosimann[1], Garrett C. Heffner[5], George Q. Daley[5], Robert F. Paulson[3,4], Richard A. Young[2] and Leonard I. Zon[1,#]*

[1] Stem Cell Program and Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA 02115, [2] Whitehead Institute for Biomedical Research, Cambridge, MA 02142, [3] Graduate program in Biochemistry, Microbiology and Molecular Biology, [4]Department of Veterinary and Biomedical Sciences at the Pennsylvania State University, University Park, PA 16802, [5] Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Children's Hospital Boston and Dana Farber Cancer Institute; Division of Hematology, Brigham and Women's Hospital; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School; Harvard Stem Cell Institute; Boston, Massachusetts 02115, USA, [6] Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02142

* These authors contributed equally

To whom correspondence should be addressed

This PDF file includes:

Supplemental data

Supplementary Figure 1 (related to Figure 1). **BMP and Wnt pathways regulate hematopoietic regeneration.** A. Schematic of murine competitive bone marrow transplantation. Limited numbers (16K) of CD45.1 bone marrow cells were treated *ex vivo* with DMSO (control), BIO (Wnt activation), or DM (BMP inhibition), then combined with 200K untreated CD45.2 congenic bone marrow cells and injected into lethally irradiated CD45.2 recipient mice. Engraftment was determined at 20 weeks post transplant. B. Graph showing the frequency of mice with greater than 1% multilineage (myeloid, T, and B cell) contribution from the CD45.1 donor treated cells. Numbers shown below each group denote the number of engrafted animals out of the total transplanted. ** Fisher's exact test p-value=0.026. C. Graphs of relative gene expression +/- SEM in kidneys isolated from wild type, Hs:Wnt, or Hs:BMP zebrafish two days post irradiation following a two hour heat shock induction of Wnt8a or BMP2b transgene expression, respectively. * Student's t-test p-values <0.05.

Supplementary Figure 2 (related to Figure 2). SMAD1 and TCF7L2 bind erythroid gene targets. A. Gene tracks representing binding of specific genomic regions along the x-axis and the total number of reads per million on the y-axis. Horizontal black bars above each track indicate the genomic scale in kilobases (kb). B. Quantile normalized composite SMAD1 enrichment profile for the SMAD1 variable regions between SMAD1 ChIPs in K562 cells stimulated with BMP4 or inhibited with dorsomorphin (DM). C. GSEA analysis of the genes bound in the top 500 enriched regions by TCF7L2 or SMAD1 in K562 cells to microarray data comparing human erythrocytes (Ery5- GlyA-positive cells) to all other cell hematopoietic cell populations. NES- normalized enrichment score, FDR q value- false discovery rate q value. D-E. D. Ingenuity Pathway Analysis (left) or E. GREAT analysis (right) of genes bound by TCF7L2 (top) or SMAD1 (bottom) in K562 cells showing a subset of enriched erythroid categories. F. Motifs enriched in TCF7L2 (top) or SMAD1 (bottom) bound regions in K562 as determined by DREME analysis. P-values listed below each motif were calculated with a Fisher's exact test. G. Histograms showing the frequency of the GATA, SMAD and TCF motifs every 250bp (y-axis) relative to the distance from the peak (x-axis).

Supplementary Figure 3 (related to Figure 2). BMP and Wnt stimulation do not affect GATA binding. GATA1 and GATA2 genome occupancy is highly overlapping in unstimulated *vs*. stimulated conditions in K562 cells. For each cell type unstimulated and stimulated (BIO or BMP4), overlaps for each GATA1 (upper) and GATA2 (lower) ChIP at each of the three different signal category (top 10,000, top 5,000 and top 1,000) cut-offs in a pictorial representation of these overlaps.

Supplementary Figure 4 (related to Figure 3). SMAD1 and TCF7L2 bind active genes. SMAD1 and TCF7L2 regions co-localize with GATA factors at the enhancers of active genes. Graph depicting the relative percentage of actively transcribed genes based on the chromatin state (H3K4me3, H3K36me3) for genes occupied by GATA and/or SMAD1, TCF7L2 compared to all RefSeq genes. p-value was calculated using a hypergeometric test to determine enrichment of active genes.

Supplementary Figure 5 (related to Figure 4). SMAD1 and TCF7L2 bind to genes that define cellular identity A. Comparison of IPA analyses of genes bound by TCF7L2 (left) or SMAD1 (right) in K562 and U937 cells. Categories listed are enriched for genes associated with either red or white blood cells. B. Comparison of GREAT analyses of regions bound by TCF7L2 (left) or SMAD1 (right) in K562 and U937 cells showing fold enrichment of listed categories above control (whole genome). Categories listed are enriched for genes associated with either red or white blood cells. C. Histograms showing the frequency of GATA, SMAD, TCF, and C/EBP motifs every 250bp (y-axis) relative to the distance from the peak (x-axis) in bound regions by GATA (includes GATA1&GATA2), SMAD1, and TCF7L2 in K562 cells (left) and C/EBPα, SMAD1 and TCF7L2 in U937 cells (right). D. C/EBPα genome occupancy is highly overlapping in unstimulated *vs*. stimulated (Wnt or BMP) conditions in U937 cells. For each cell type unstimulated and stimulated (BIO or BMP4), overlaps for each C/EBPα ChIP at each of the three different signal category (top 10,000, top 5,000 and top 1,000) cut-offs in a pictorial representation of these overlaps.

Supplementary Figure 6 (related to Figure 6). SMAD1 binding becomes lineage-restricted during erythroid differentiation A. Histogram showing frequency of SMAD and GATA motifs every 250bp (y-axis) relative to the distance from the peak (x-axis) in the bound regions by GATA2 and SMAD1 in G1E and SMAD1 and GATA1 in G1ER cells. B. Comparison of IPA analyses of genes bound by SMAD1 in G1E proerythroblasts and G1ER differentiated erythroid cells. Categories listed are enriched for genes associated with either red or white blood cells. C. Comparison of GREAT analyses of regions bound by SMAD1 in G1E proerythroblasts and G1ER differentiated erythroid cells showing fold enrichment of listed categories above control (whole genome). Categories listed are enriched for genes associated with either red or white blood cells.

Supplementary Figure 7 (related to Figure 7). SMAD1 and TCF7L2 occupied regions change during differentiation. A**.** Flow cytometry plots showing the expression of progenitor and differentiation markers at days 0, 1, and 5 during erythroid differentiation. Left plots show expression of CD34 and CD38 on day 0 of differentiation. Top plot is a negative control and bottom plot shows 83% of cells after expansion remain CD34+. Middle plots show expression of CD34/CD36 (top) or CD71/GlycophorinA on day 1 of differentiation and right plots show the same for day 5 of differentiation. B. Region plots representing the distribution of GATA2 bound regions -2.5 to +2.5 kb relative to all TCF7L2 (left) or SMAD1 (right) bound regions in CD34+ progenitors. C. IPA (left) and GREAT (right) analyses of genes bound by TCF7L2 in CD34 progenitors (CD34pro). Categories listed are enriched for genes associated with either red or white blood cells. D. IPA (left) and GREAT (right) analyses of genes bound by SMAD1 in CD34 progenitors (CD34pro) and CD34-derived erythroblasts (CD34ery). Categories listed are

enriched for genes associated with either red or white blood cells. IPA analyses for S7C and S7D were performed using the genes associated with the top 1000 bound regions. E. Histograms showing the frequency of SMAD, TCF, ETS, RUNX, and GATA motifs every 250bp (y-axis) relative to distance from the peak (x-axis) in bound regions by GATA2, SMAD1, and TCF7L2 in CD34+ progenitors and SMAD1 and GATA1 in CD34-derived erythroblasts.

Supplemental Experimental Procedures

Zebrafish Irradiation-induced Regeneration

Zebrafish were maintained as described (Westerfield, 2000). All fish were maintained according to IACUC approved protocols in accordance with Children's Hospital Boston animal research guidelines. Irradiation-induced regeneration assays were performed as previously described (Burns et al., 2005; Traver et al., 2004). For heat shock treatments, small fish tanks were placed at 37°C for 16 hours. The transgenic zebrafish lines Hs:Wnt8, Hs:Dkk1, Hs:BMP, and Hs:Chordin were described previously (Rentzsch et al., 2006; Stoick-Cooper et al., 2007; Tucker et al., 2008; Weidinger et al., 2005). For small molecule treatments, fish were soaked for 16 hours in fish water containing drug dissolved in DMSO, 2μM BIO or 10μM DM. On days 2, 7, and 14 post irradiation, whole kidney marrow cells were analyzed on an LSRII for FSC/SSC parameters.

Mouse Bone Marrow Transplantation

All mice were maintained according to IACUC approved protocols in accordance with Children's Hospital Boston animal research guidelines. All mouse strains were purchased from Jackson laboratories and housed in autoclaved cages with autoclaved water. Competitive transplantation was performed as previously described with the following modifications (Harrison, 1980). Briefly, 9 week old CD45.2 (C57/Bl6) recipient mice were irradiated with an 11Gy split dose of γ-irradiation. Bone marrow cells (16K) from 9 week old CD45.1 mice were treated *ex vivo* with DMSO, 0.5μM BIO, or 5μM DM for 3 hours at 37°C. Cells were then combined with a 12.5-fold excess of bone marrow cells (200K) from CD45.2 mice and transplanted into CD45.2 recipients. For each treatment 4-8 mice were transplanted per

experiment with three biological repeats. Peripheral blood was collected by retro-orbital bleeding at 6, 12, and 20 weeks post transplant. Samples were stained and then analyzed on an LSRII (BD Biosciences) to assess multilineage engraftment. Animals with 1% chimerism in myeloid, T, and B cell lineages were considered to have multilineage engraftment. CD45.1 chimerism for each recipient mouse is shown in Supplementary Table 1.

RT-PCR analysis of zebrafish WKM

Gene expression analysis on zebrafish WKM following irradiation was done as previously described with the following modification (Burns et al., 2005). Single transgenic Hs:Wnt8a, Hs:BMP2b, and their wild type siblings were irradiated on day 0, incubated at 37°C for 2 hours on day 2 following irradiation, and then euthanized for cell isolation. Kidneys from zebrafish were dissected and placed in Trizol (Invitrogen 15596-026) for RNA isolation according to manufacturer's instructions. Contaminating DNA was removed using DNA-free (Ambion AM1906) and cDNA was generated using Superscript III Supermix (Invitrogen 11752). Three-six biological replicates and technical duplicates were used for each genotype. QPCR reactions were done with iQ SYBR-GREEN Supermix (BIORAD 170-8880), and a CFX-384 Real-Time PCR Detection System (BIORAD) thermal cycler was used. Expression of each gene was normalized to beta-actin and gapdh and relative levels were calculated using the $\Delta\Delta$Ct method (Applied Biosystems). Quantities are expressed as fold change compared to wild type controls. Primers are listed in Supplementary Table 2.

Mouse sublethal irradiation injury and isolation of lineage negative bone marrow cells

For irradiation injury, 9-week old C57/Bl6 mice were irradiated with a 6.5Gy dose of γ-irradiation as previously described (Hooper et al., 2009). Bone marrow was isolated from 20 mice on day 7 following irradiation. Bones were crushed to release bone marrow cells and then filtered through a 40μm cell strainer. Red blood cells were lysed using an ammonium chloride solution, and the remaining cells were then incubated for 30 min at 4°C with rat anti-mouse lineage antibodies for T cells (CD4-BD Biosciences 553727 and CD8-BD Biosciences 553027), B cells (B220-Invitrogen RM2600), myeloid cells (Mac1-Invitrogen RM2800 and GR1-Invitrogen RM3000), and erythrocytes (TER119-BD Biosciences 550565). After washing, the cells were incubated for 30 min at 4°C with sheep anti-rat magnetic dynabeads. The cell-dynabead mixture was placed on a dynal magnet to remove the lineage-positive cells. The lineage-negative cells from the remaining supernatant were crosslinked for ChIP as described below.

Cell line culture conditions and stimulation

K562 and K562 overexpressing C/EBPα-ER (K562-C/EBPα) cells were maintained in IMDM medium (Invitrogen 12440) with 10% Fetal Bovine serum, 2μM Glutamine, 100 U/mL penicillin and 100 μg/mL streptomycin (1% P/S). Cells were split 1:10 every 2-3 days. Prior to stimulation, cells were serum starved in IMDM medium with 2μM Glutamine, 1% P/S for 24hrs at a concentration of 10^6cells/ml. For TCF7L2 ChIP-seq, cells were serum starved for 24 hrs and stimulated with 5μM BIO dissolved in DMSO for 4hrs. For the SMAD1 ChIP-seq, cells were serum starved for 24 hrs and incubated with 25ng/μl recombinant human BMP4 (R&D 314-BP-

010) for 2hrs or 20μM dorsomorphin (Calbiochem AMPK inhibitor, Compound C 171260) for 2hrs. For GATA1 and GATA2 ChIP-seq, cells were serum starved for 24 hrs, and then treated with BIO, rhBMP4, or vehicle control. Prior to stimulation K562-C/EBPα cells were serum starved for 24 hrs; at the same time estradiol was added to a final concentration of 200nM. For the SMAD1 and C/EBPα ChIP-seq, cells were stimulated with 25ng/μl recombinant human BMP4, for the last 2hrs of the 24hrs induction with estradiol.

U937 cells were maintained in RPMI medium (Invitrogen 12633) with 10% Fetal Bovine serum, 2μM Glutamine, 1% P/S. Cells were split 1:10 every 2-3 days. Prior to stimulation, cells were serum starved in RPMI medium with 2μM Glutamine, 1%P/S for 24hrs at a concentration of 10^6cells/ml. Cells were stimulated with BIO and rhBMP4 the same as for K562, except BIO stimulation was for 2 hours instead of 4 hours. For the C/EBPα ChIP-seq, cells were serum starved for 24 hrs, and then treated with BIO, rhBMP4, or vehicle control.

Human $CD34^+$ cells, isolated from the peripheral blood of granulocyte colony-stimulating factor mobilized healthy volunteers, were obtained from the Yale Center of Excellence in Molecular Hematology. The cells were maintained and differentiated as previously described (Sankaran et al., 2008). Briefly the cells were expanded in StemSpan medium (Stem Cell Technologies Inc.) supplemented with 1X CC100 cytokine mix (Stem Cell Technologies Inc.) and 2% P/S for a total of 6 days. At day 6 cells were stimulated for 2hrs with rhBMP4 at a final concentration of 25ng/μl or 5μM of BIO as indicated and harvested for performing chromatin immunoprecipitation. For studying the binding in differentiated cells after day 6 of expansion, cells were reseeded in differentiation medium (StemSpan SFEM Medium with 2% P/S, 20 ng/ml

SCF, 1 U/ml Epo, 5 ng/ml IL-3, 2 uM dexamethasone, and 1 uM β-estradiol), at a density of 0.5–1 X 10^6 cells/ml till harvesting. At day 5 of differentiation, cells reached the proerythroblast stage of erythroid differentiation. At this day cells were stimulated for 2hrs with rhBMP4 at a final concentration of 25ng/µl and harvested for performing ChIP. Flow cytometry on the cells at days 0, 1, and 5 during differentiation was performed to determine the hematopoietic phenotype of the cells. CD34-PECy5 (BD Biosciences 555823) and CD38-PE (BD Biosciences 555460) were used to examine progenitors at day 0 of erythroid differentiation, which was on the sixth day of expansion. CD71-APC (BD Biosciences 551374), GlycophorinA-PE (BD Biosciences 555570), and CD36-FITC (BD Biosciences 555454) were used to determine the stage of erythroid differentiation on days 1 and 5 of differentiation.

G1E and G1ER cells were maintained in IMDM medium plus 15% heat-inactivated fetal calf serum (Hyclone SH30071.01) in the presence of 2% P/S, 2 U/mL erythropoietin, 50ng/ml mouse stem-cell factor (R&D 455-MC-010), and 124 X 10^{-4} monothioglycerol (Sigma M6145) as previously described (Tsang et al., 1997). Erythroid differentiation was induced with 10^{-7} M β-estradiol for 24hrs both in G1ER but also in the G1E cells as a control. During the last two hours, rhBMP4 was added to the cultures at a final concentration of 25ng/µl and cells were harvested for immunoprecipitation.

Reporter assays

K562 cells were transfected using AMAXA nucleofector with plasmids encoding FLAG-GATA2, MYC-SMAD1, GFP-TCF7L2, -72 LMO2 enhancer conjugated to LMO2 promoter LacZ (Landry et al., 2009) in combinations, as depicted in Figure 2E, according to the manufacturer's

instructions. A plasmid encoding *Renilla* luciferase was used as a control for transfection efficiency. β-galactosidase activity was measured with the Galacto-Star kit (Applied Biosystems) and *Renilla* luciferase was measured with the *Renilla* Luciferase Assay System (Promega).

Chromatin Immunoprecipitation

A summary of the bound regions and bound genes determined for all ChIP-seq data is contained within Supplementary Tables 3,4,5.

For ChIP-seq experiments the following antibodies were used: Smad1 (Santa Cruz sc7965), TCF7L2 (TCF4 Santa Cruz sc8631), C/EBPα (Santa Cruz sc9314), Gata1 (Santa Cruz sc265), Gata2 (Santa Cruz sc9008), p300 (Santa Cruz sc585), and H3K4me1 (Abcam ab8895).

Chromatin immunoprecipitation experiments were performed as previously described (Lee et al., 2006) with slight modifications. Briefly, 10^8 cells for K562 and U937 cell lines and 1-5x10^7 cells for other cells used, were crosslinked by the addition of 1/10 volume 11% fresh formaldehyde for 10 min at room temperature. The crosslinking was quenched by the addition of 1/20 volume 2.5M Glycine. Cells were washed twice with ice-cold PBS and the pellet was flash-frozen in liquid nitrogen. Cells were kept at -80^{o}C until the experiments were performed. Cells were lysed in 10ml of Lysis buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1× protease inhibitors) for 10min at 4^{o}C. After centrifugation, cells were resuspended in 10 ml of Lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors) for 10 min at room

temperature. Cells were pelleted and resuspended in 3ml of Sonication buffer for K562 and U937 and 1ml for other cells used (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.05% *N*-lauroylsarcosine, 1× protease Inhibitors) and sonicated in Bioruptor sonicator for 24 cycles of 30sec of sonication followed by 1min resting intervals. Samples were centrifuged for 10min at 18,000g and 1% of Triton X was added to the supernatant. Approximately 1-3x10^7 cells were used for each immunoprecipitation. Prior to the immunoprecipitation, 50 µl of protein G beads (Invitrogen 100-04D) for each reaction were washed twice with PBS, 0.5% BSA twice. Finally the beads were resuspended in 250µl of PBS, 0.5% BSA and 5µg of each antibody. Beads were rotated for at least 6hrs at 4°C and then washed twice with PBS, 0.5% BSA. Cell lysates were added to the beads that were incubated at 4°C overnight. Beads were washed 1X with 20mM Tris-HCl pH8, 150 mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10mM Tris-HCl pH8, 250nM LiCl, 2mM EDTA, 1% NP40 and 1X with TE and finally resuspended in 200µl elution buffer (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 0.5-1% SDS) by heating at 65°C for 30 minutes in a shaking heat block. 50µl of cell lysates prior to addition to the beads was kept as input. Crosslinking was reversed by incubating samples at 65°C for at least 6 hours. After the reversal of crosslinking cells were treated with RNase and proteinase K and the DNA was extracted by Phenol/Chloroform extraction.

Sequential Chromatin immunoprecipitation

ChIPs were performed as described above until the point of the reversal of crosslinking/bead elution. At that point the immunoprecipitated DNA fragments were eluted from the beads by

addition of 55µl 10mM DTT and incubated at 37^{o}C for 30 min. The supernatant was transferred to a new tube and the material was diluted 30X with sonication buffer that contained 1% Triton X. This dilution served as the starting material for the second ChIP. The rest of the ChIP was performed as described above. The genes tested are blood gene targets that were bound by both GATA2 and SMAD1 in the ChIP-Seq data. QPCR reactions were done with iQ SYBR-GREEN Supermix (BIORAD 170-8880), and a CFX-384 Real-Time PCR Detection System (BIORAD) thermal cycler was used. Expression of each gene was normalized to beta-actin and relative levels were calculated using the $\Delta\Delta$Ct method (Applied Biosystems). Quantities are expressed as fold change compared to whole cell extract (input) controls. Primers are listed on Supplementary Table 2.

ChIP-seq sample preparation and analysis

All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203).

Sample preparation

All samples were prepared with the Illumina/Solexa Genomic DNA kit (Illumina- IP-102-1001) according to the manufacturer's instructions. Briefly 200ng of Input DNA and various amounts of ChIP DNA were used. DNA overhangs were turned into phosphorylated blunt ends and the samples were purified with the PCR purification kit (Qiagen 28104). Sample preparation continued by the addition of a single A in the 3' end to allow for directional ligation. Samples were purified with the MinElute PCR purification kit (Qiagen 28004). Illumina adapter oligos (1/100 dilution) were added to the sample followed by purification of the samples with the PCR

purification kit. The samples were amplified by PCR (limited amplification to 18 cycles) that added additional linker sequence to the fragments to prepare them for annealing to the Genome Analyzer flow-cell. The amplified samples were separated on a 2% agarose gel (products between 150-350 base pairs were selected that include fragments of 50-250bp with approximately 100bp of primer sequence) and gel extraction was performed with the Gel Extraction Kit (Qiagen 28704).

For multiplexed samples, libraries were prepared using the Illumina TruSeq adapters (to enable multiplexing), and prepared using Beckman-Coulter's SPRIworks system. ChIP samples were used in their entirety as input, and whole cell extracts were prepared using 100 ng of input. Adapters were diluted to 1:200. Size selection was 200-400 bp before PCR, and the samples were amplified using KAPA Hi-Fi polymerase and 18 cycles of PCR according to manufacturer's cycling recommendations. PCR products were purified using Agencourt Ampure XP beads using a 0.9X ratio of beads to sample.

Polony generation and sequencing

The DNA library (2-5 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1 µm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent

illumination. Validated flow-cells were stored at 4^{o}C until sequencing. Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer II or Hi-seq. After the first base was incorporated in the Sequencing by-Synthesis reaction the process was paused for a quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were reprimed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26 cycles. For multiplexed samples Truseq V2.5 kits were used to cluster them on the cBot and Truseq V2 were used to do a multiplex 40+7 cycle run on the Hi-seq.

ChIP-seq data analysis

Images acquired from the Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which identified polony positions, performed base-calling and generated QC statistics. Sequences were aligned using ELAND software to NCBI Build 36 (UCSC hg18) of the human genome and NCBI Build 36 (UCSC mm8) of the mouse genome. Only sequences that mapped uniquely to the genome with zero or one mismatch were used for further analysis. When multiple reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used to determine enriched regions as described below.

Analysis methods were derived from previously published methods (Guenther et al., 2008; Johnson et al., 2007; Marson et al., 2008; Mikkelsen et al., 2007). Each read was extended

200bp except H3K36me3 extended 1 Kbp, towards the interior of the sequenced fragment, based on the strand of the alignment. Across the genome the number of ChIP-seq reads was tabulated in 10 bp bins for transcription factors, 25 bp bins for histone marks. The genomic bins that contained statistically significant ChIP-seq enrichment were identified by comparison to a Poissonian background model. Assuming background reads are spread randomly throughout the genome, the probability of observing a given number of reads in a 1kb window can be modeled as a Poisson process in which the expectation can be estimated as the number of mapped reads multiplied by the number of bins into which each read maps, divided by the total number of bins available. Enriched bins within 200bp of one another were combined into regions. The Poissonian background model assumes a random distribution of background reads. However, significant deviations from this expectation have been observed. Some of these non-random events can be detected as sites of apparent enrichment in negative control DNA samples creating false positives. To remove these false positive regions, negative control DNA from whole cell extract (WCE) was sequenced for each cell type. Enriched bins and enriched regions were defined as having greater than five-fold density in the experimental sample compared with the control sample when normalized to the total number of reads in each dataset. This served to filter out genomic regions that are biased to having a greater than expected background density of ChIP-seq reads. For human, the complete set of RefSeq genes was downloaded from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables?command=start) on March 1, 2009. For mouse, the complete set of RefSeq genes was downloaded from the UCSC website

(http://hgdownload.cse.ucsc.edu/goldenPath/mm8/database/) on March 5, 2010. Enriched regions within 5 kb upstream or downstream of the body of the gene were called bound.

Lists of the bound regions and bound genes (Supplementary Tables 4-5) for each factor is provided in a separate sheet, and the summary of bound regions and bound genes are provided in Supplementary Table 3. Additionally, data files that contain genome browser tracks showing genome-wide ChIP-seq density and enriched regions for all experiments are available on GEO under the Superseries accession number GSE29196 with datasets under accession numbers GSE29193, GSE29194, and GSE29195.

Previously published ChIP-seq quality score (FASTQ) files profiling the genomic occupancy of GATA1, GATA2, MYC, MAX, E2F4, CTCF, YY1, cFos, H3K4me3, H3K36me3, and H3K27me3 in K562 cells were downloaded from UCSC ENCODE website (http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg18&g= wgEncodeBroadChipSeq). Briefly Bowtie (version 0.12.2)(Langmead et al., 2009) was used to align sequences to UCSC human genome build hg18. Alignments were performed using the following criteria: -n2, -e70, -m2, -k2, --best.

Relative Change in SMAD1 binding in Stimulated and Inhibited K562

For comparison of SMAD1 occupancy in K562 cells following BMP4 stimulation or Dorsomorphin inhibition, rank-based quantile normalization was used to normalize the datasets to be compared (Guenther et al., 2010). Briefly, for each dataset compared, the genomic bin with the greatest ChIP-Seq density was identified. The mean of these values was calculated and

the bin with the greatest signal in each sample was assigned this mean value. This was repeated for all genomic bins from the greatest signal to the least, assigning each the average ChIP-Seq signal for all bins of that rank across all datasets.

GREAT analysis

The enriched regions from each ChIP-Seq were imported into Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean, Bejerano Nat biotech 2010) from the link from UCSC genome tables. GREAT associates genomic regions with putative target genes, then uses gene annotations from numerous ontologies including gene ontology (GO), mouse phenotype, and disease ontology and then calculates statistical enrichment of each category in enriched regions compared to the whole genome. We considered categories with p-values < 10^{-4} as significant.

Ingenuity pathway analysis

The enriched genes from each ChIP-Seq were imported into Ingenuity Pathways Analysis (IPA) (Ingenuity® Systems, http://www.ingenuity.com) to analyze functional interactions between the genes. For analyses of genes bound in CD34 cells, the genes associated with the top 1000 bound regions were assessed. The Functional Analysis identified the biological functions and/or diseases that were most significant to the dataset. Molecules from the dataset associated with biological functions and/or diseases in Ingenuity's Knowledge Base were considered for the analysis. Right-tailed Fisher's exact test was used to calculate a p-value determining the

probability that each biological function and/or disease assigned to that data set is due to chance alone. The applied threshold was of q-value of <0.05.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was performed as previously described (Mootha et al., 2003; Subramanian et al., 2005). Transcription factor bound gene lists were compared to microarray data comparing sorted human hematopoietic populations including hematopoietic stem cells, nucleated erythrocytes, megakaryocytes, monocytes, granulocytes, dendritic cells, NK, T, and B cells (Novershtern et al., 2011). Default parameters for GSEA were used including a weighted enrichment statistic, Signal2Noise metric for ranking genes, real gene list sorting mode, and descending gene list ordering mode. One thousand permutations were run on the phenotype parameter for each comparison. For analyses of genes bound by TCF7L2 or SMAD1 in K562, the top 500 genes were used.

Active, Initiated, Bivalent and Silent Gene Classes in K562 Cells

The active and initiated genes were classified in K562 cells using H3K4me3 (initiation-associated chromatin modification) and H3K36me3 (elongation-associated chromatin modification), as determined by ChIP-seq in K562 cells, as markers of transcriptional state. Active genes with both H3K4me3 and H3K36me3 chromatin modifications, initiated genes had only H3K4me3, bivalent genes contained H3K4me3 and H3K27me3 and silent genes did not have H3K4me3 or H3K36me3 chromatin modifications.

Genome-wide expression analysis

Affymetrix U133plus2.0 microarrays were used to assess gene expression changes in CD34+ cells with or without Wnt or BMP pathway stimulation. Arrays were done on control, 0.5μM BIO, or 25ng/ml rhBMP4 treated cells two hours after a shift to erythroid differentiation media. Raw expression data was preprocessed by MAS5 and then log10 transformed. Biological replicates across arrays were then quantile normalized. We used the probe with the maximum expression change as the representative for that gene. A gene was called to be significantly differentially expressed if it meets the following criteria: 1) at least a 1.5 fold difference (absolute log10 difference of greater than 0.176) between mean expression at 2 hour after differentiation treatment, 2) P value smaller than 0.01 by SNR test. The enrichment of bound genes in the differentially expressed genes was calculated using a Chi-square test. Lists of differentially expressed genes are in Supplementary Table 6.

Supplemental References

Burns, C.E., Traver, D., Mayhall, E., Shepard, J.L., and Zon, L.I. (2005). Hematopoietic stem cell fate is established by the Notch-Runx pathway. Genes Dev *19*, 2331-2342.

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249-257.

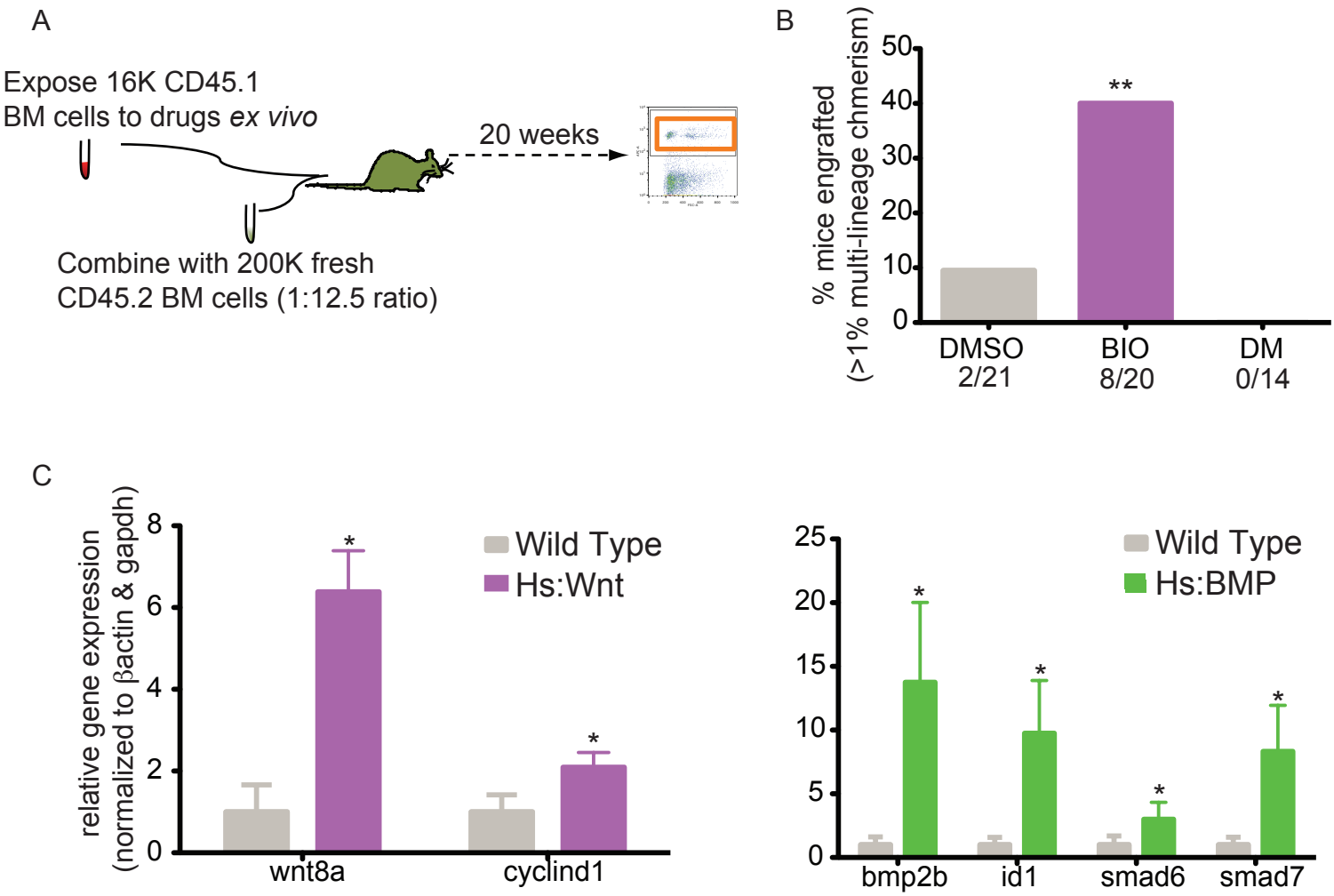
Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E., and Young, R.A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev *22*, 3403-3408.

Harrison, D.E. (1980). Competitive repopulation: a new assay for long-term stem cell functional capacity. Blood *55*, 77-81.

Hooper, A.T., Butler, J.M., Nolan, D.J., Kranz, A., Iida, K., Kobayashi, M., Kopp, H.G., Shido, K., Petit, I., Yanger, K.*, et al.* (2009). Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. Cell Stem Cell *4*, 263-274.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.
Landry, J.R., Bonadies, N., Kinston, S., Knezevic, K., Wilson, N.K., Oram, S.H., Janes, M., Piltz, S., Hammett, M., Carter, J.*, et al.* (2009). Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. Blood *113*, 5783-5792.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.
Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.
Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E.*, et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet *34*, 267-273.
Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T.*, et al.* (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296-309.
Rentzsch, F., Zhang, J., Kramer, C., Sebald, W., and Hammerschmidt, M. (2006). Crossveinless 2 is an essential positive feedback regulator of Bmp signaling during zebrafish gastrulation. Development *133*, 801-811.
Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839-1842.
Stoick-Cooper, C.L., Weidinger, G., Riehle, K.J., Hubbert, C., Major, M.B., Fausto, N., and Moon, R.T. (2007). Distinct Wnt signaling pathways have opposing roles in appendage regeneration. Development *134*, 479-489.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Traver, D., Winzeler, A., Stern, H.M., Mayhall, E.A., Langenau, D.M., Kutok, J.L., Look, A.T., and Zon, L.I. (2004). Effects of lethal irradiation in zebrafish and rescue by hematopoietic cell transplantation. Blood *104*, 1298-1305.
Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M., and Orkin, S.H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell *90*, 109-119.
Tucker, J.A., Mintzer, K.A., and Mullins, M.C. (2008). The BMP signaling gradient patterns dorsoventral tissues in a temporally progressive manner along the anteroposterior axis. Dev Cell *14*, 108-119.
Weidinger, G., Thorpe, C.J., Wuennenberg-Stapleton, K., Ngai, J., and Moon, R.T. (2005). The Sp1-related transcription factors sp5 and sp5-like act downstream of Wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. Curr Biol *15*, 489-500.
Westerfield, M. (2000). The zebrafish book : a guide for the laboratory use of zebrafish (Danio rerio), Ed. 4. edn (Eugene, Or., M. Westerfield), pp. 1-35.

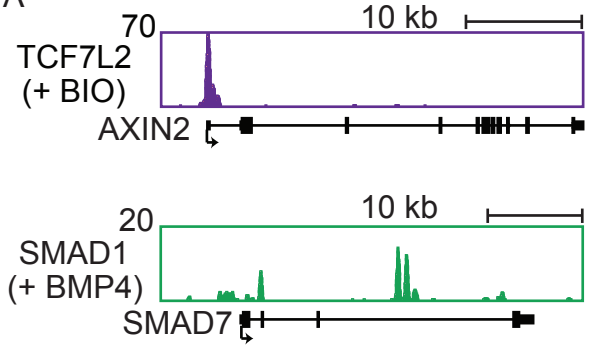
Supplemental Figure 1



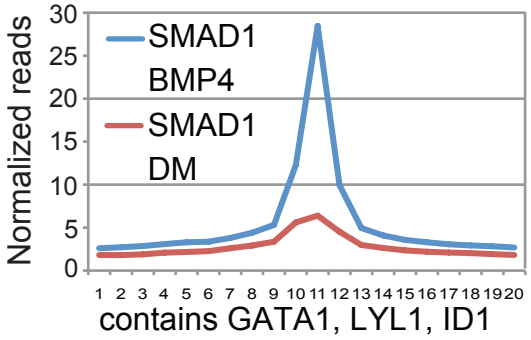
Trompouki, Bowman et al. Figure S1

Supplemental Figure 2

A

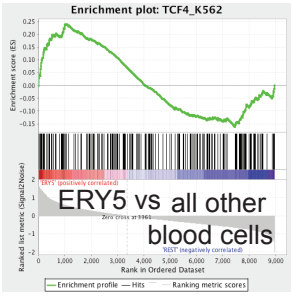


B



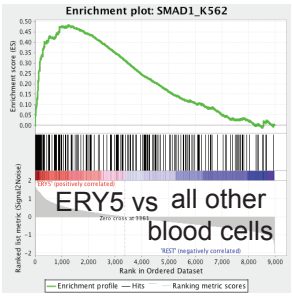
C

TCF7L2



NES 1.33 fdr q=0.014

SMAD1



NES 2.62 fdr q=0

D IPA Analysis

TCF7L2-IPA	p-values
Differentiation of red blood cells	2.9E-05
Differentiation of erythroblasts	1.1E-03

SMAD1-IPA	p-values
Differentiation of red blood cells	1.6E-08
Differentiation of erythroblasts	1.0E-04

E GREAT Analysis

TCF7L2-GREAT	Fold enrichment
JAK-STAT signaling	2.49
Chronic myeloid leukemia	2.52
Neighborhood of TAL1	2.77

SMAD1-GREAT	Fold enrichment
Erythrocyte differentiation	2.19
Anemia	2.14
Neighborhood of TAL1	2.83

F

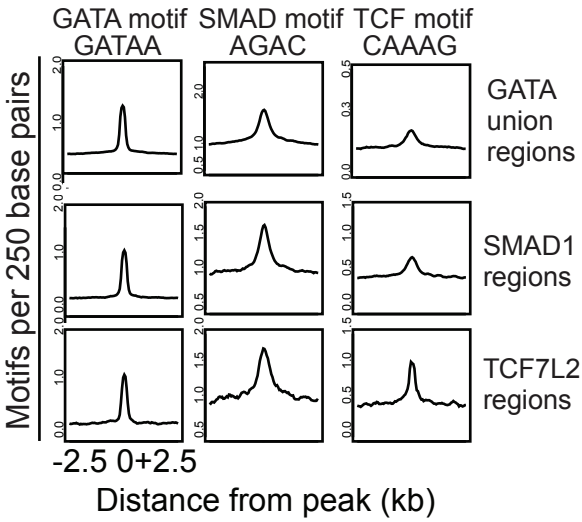
TCF7L2

GATA (4.8E-189) RUNX (9.1E-37) TCF (5.5E-51)

SMAD1

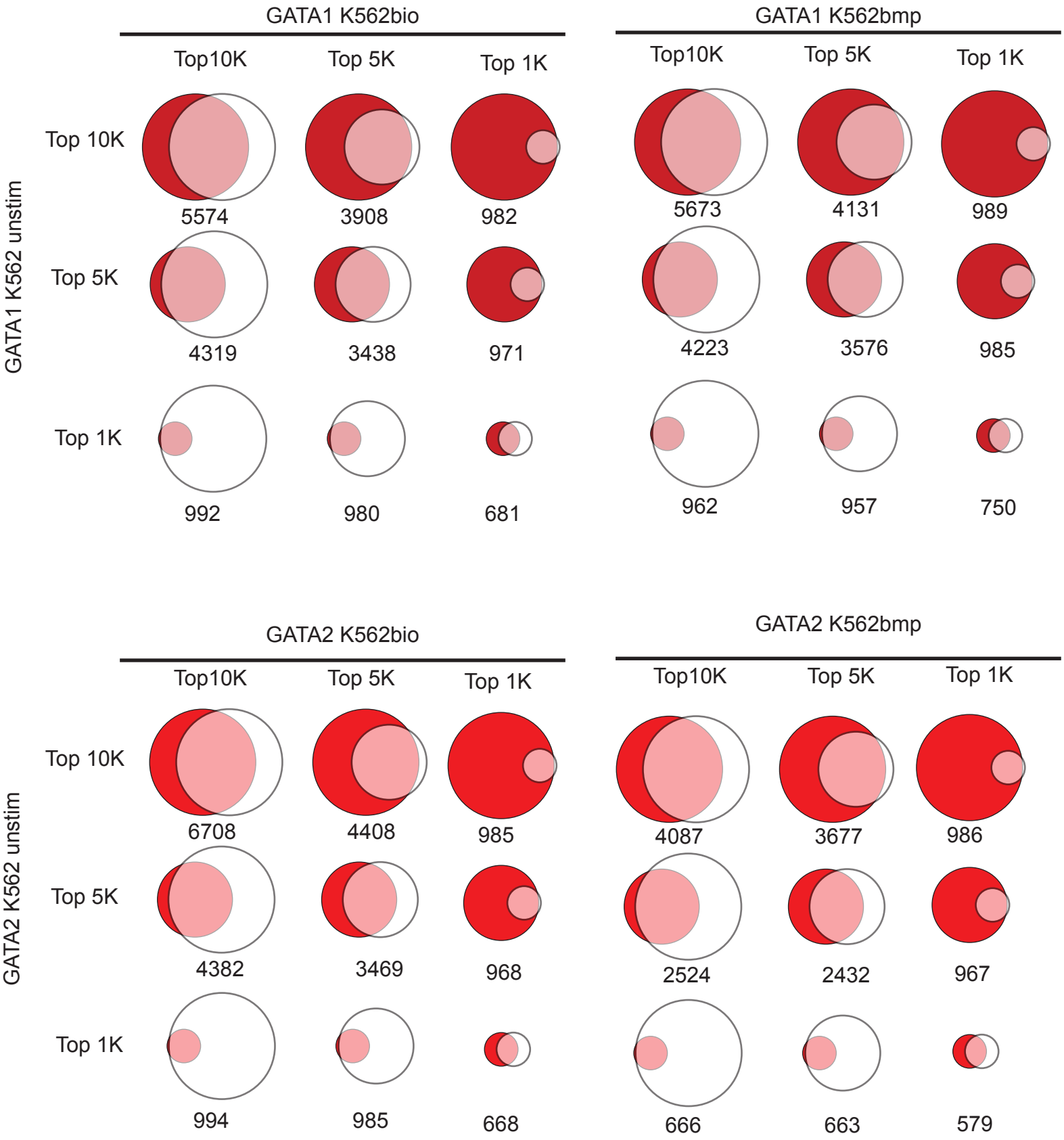
GATA (1.4E-1087) RUNX (5.4E-78) E-BOX (2.8E-51)

G



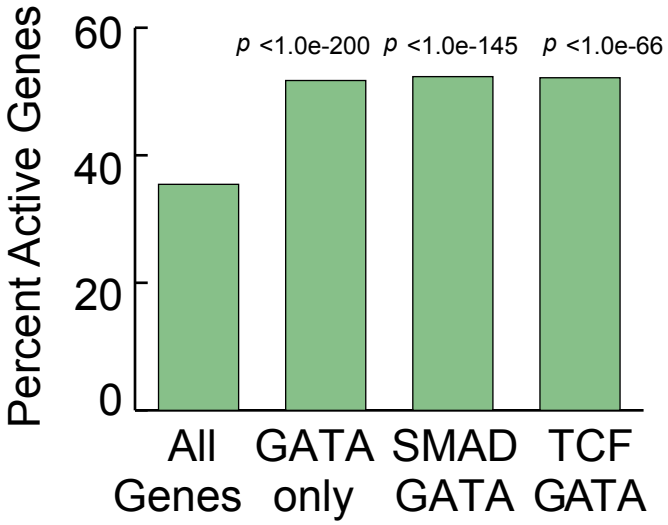
Trompouki, Bowman et al. Figure S2

Supplemental Figure 3



Trompouki, Bowman et al. Figure S3

Supplemental Figure 4



Trompouki, Bowman et al. Figure S4

Supplemental Figure 5

A

TCF7L2-IPA	p-values-K562 cells	p-values-U937 cells
Differentiation of erythroblasts	1.1E-03	Not enriched
Differentiation of leukocytes	9.7E-04	3.7E-10
Development of lymphocytes	2.9E-06	1.4E-09

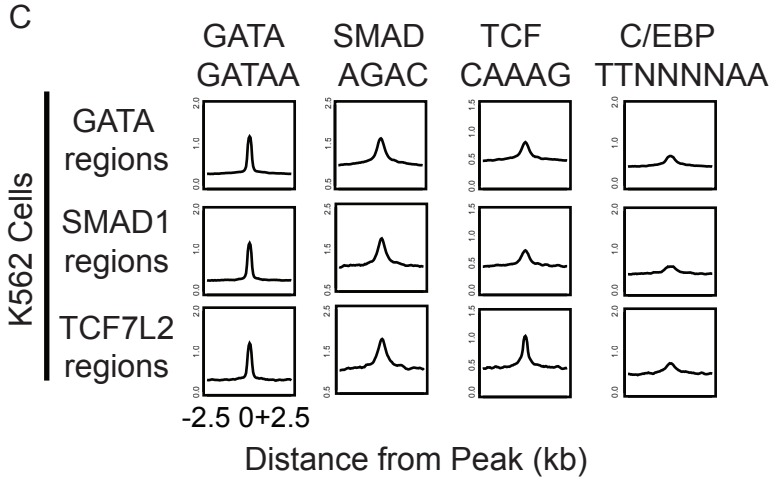
SMAD1-IPA	p-values-K562 cells	p-values-U937 cells
Differentiation of red blood cells	1.6E-08	Not enriched
Differentiation of erythroblasts	1.0E-04	Not enriched
Differentiation of leukocytes	1.4E-07	7.8E-19
Development of lymphocytes	8.0E-10	8.6E-22

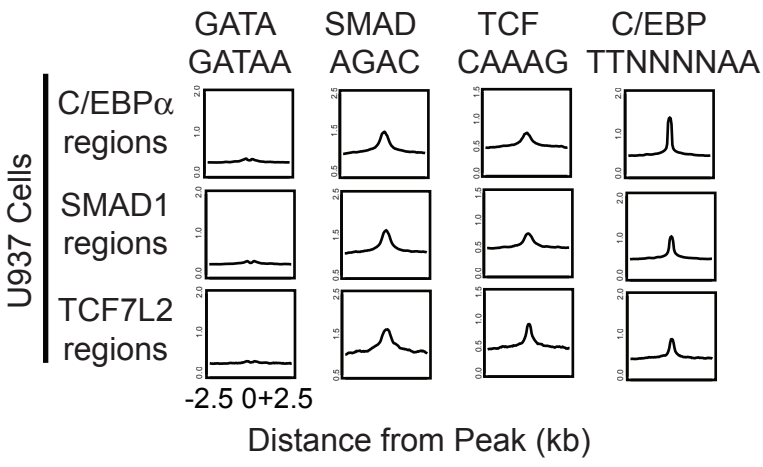
B

TCF7L2-GREAT	Fold enrichment-K562 cells	Fold enrichment-U937 cells
JAK-STAT signaling	2.49	Not enriched
Chronic myeloid leukemia	2.52	Not enriched
Neighborhood of TAL1	2.77	Not enriched
Positive regulation of myeloid cells	Not enriched	2.8
Abnormal T cell physiology	Not enriched	2.02
Cell chemotaxis	Not enriched	2.51

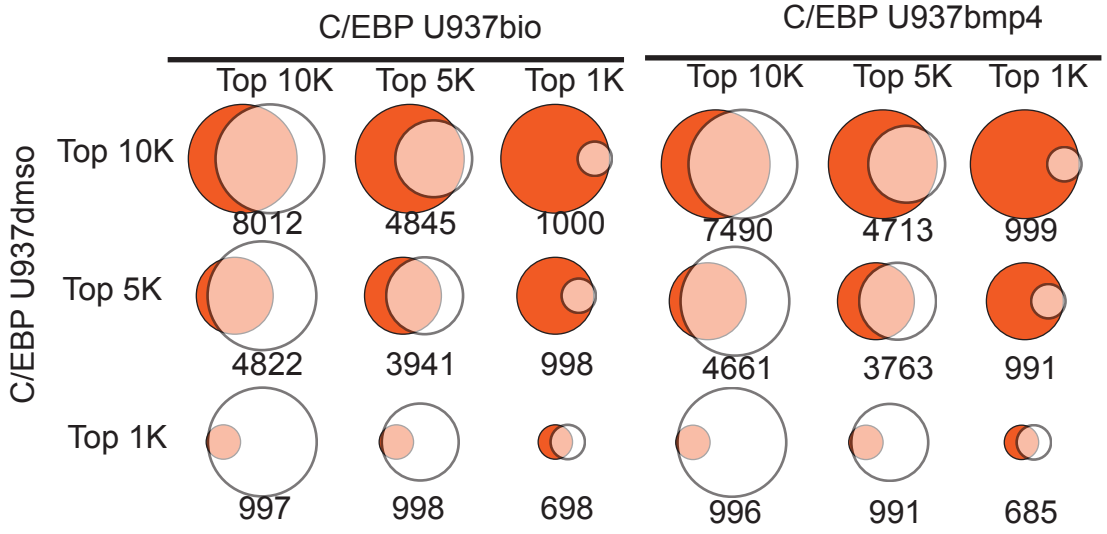
SMAD1-GREAT	Fold enrichment-K562 cells	Fold enrichment-U937 cells
Erythrocyte differentiation	2.19	Not enriched
Anemia	2.14	Not enriched
Neighborhood of TAL1	2.83	Not enriched
Positive regulation of myeloid cells	Not enriched	2.83
Abnormal T cell physiology	Not enriched	2.11
Cell chemotaxis	Not enriched	2.46

C





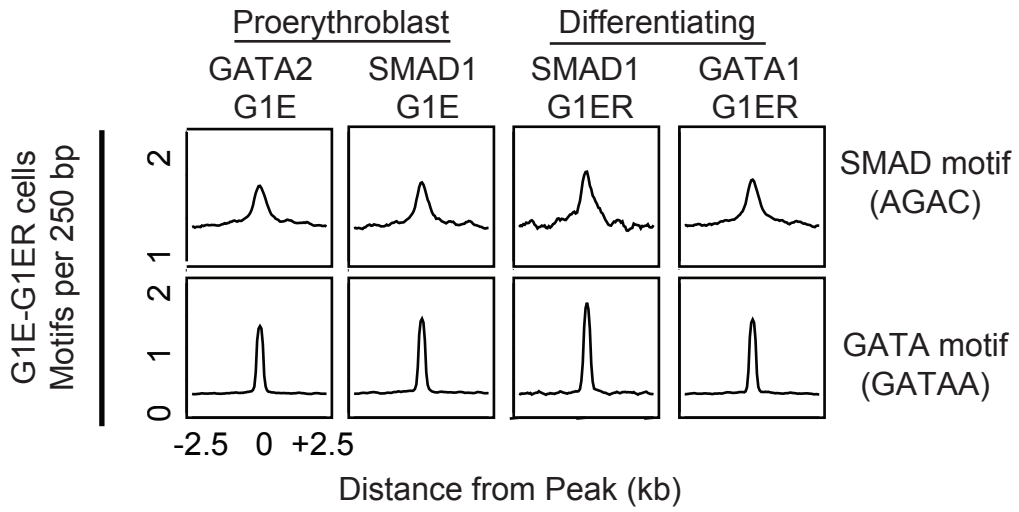
D



Trompouki, Bowman et al, Figure S5

Supplemental Figure 6

A



B

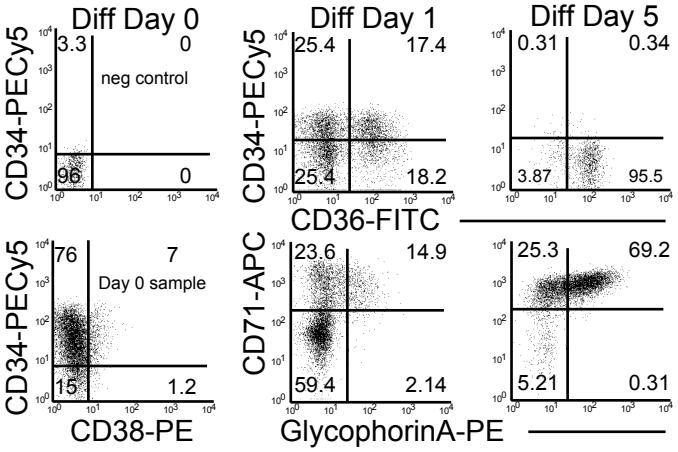
SMAD1-IPA	p-values-G1E	p-values-G1ER	p-values-G1E NOT IN G1ER
Development of leukocytes	2.3E-10	Not enriched	1.3E-06
Development of lymphocytes	5.9E-10	5.3E-05	3.2E-06
Development of red blood cells	4.0E-05	9.8E-06	Not enriched
Development of erythroid precursor cells	Not enriched	1.9E-03	Not enriched

C

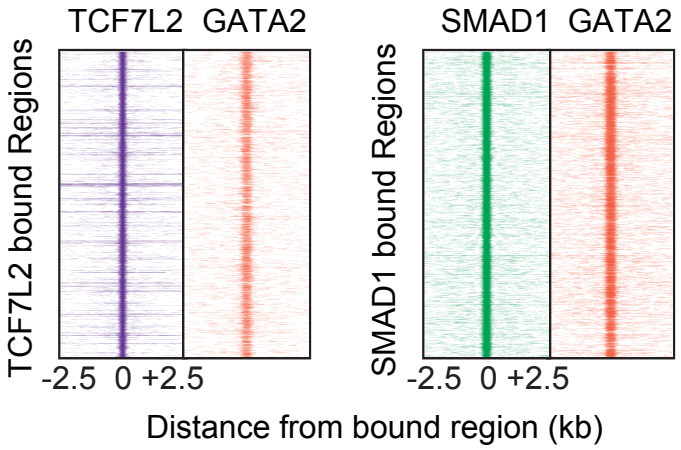
SMAD1-GREAT	Fold enrichment-G1E	Fold enrichment-G1ER
Erythrocyte differentiation	2.9	3.0
Anemia	2.2	2.7
Regulation of T cell proliferation	2.0	Not enriched
Negative regulation of lymphocyte activation	2.2	Not enriched

Supplemental Figure 7

A



B



C

TCF7L2-Pro-IPA	p-values
Differentiation of red blood cells	Not enriched
Differentiation of erythroblasts	Not enriched
Differentiation of leukocytes	1.3E-02
Development of lymphocytes	7.0E-03

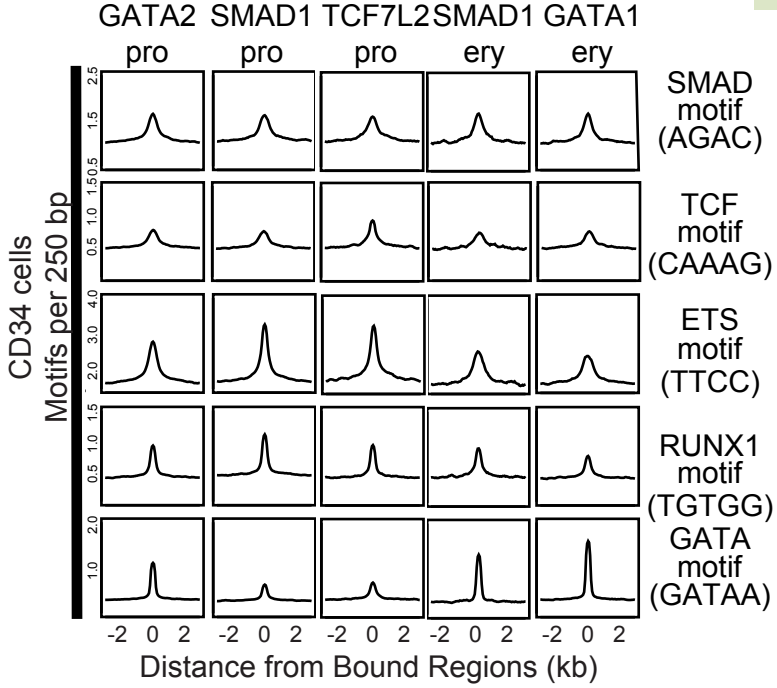
TCF7L2-Pro-GREAT	Fold enrichment
Regulation of lymphocyte differentiation	2.49
Positive regulation of myeloid cell differentiation	2.52
EPO signaling pathway	2.77

D

SMAD1-IPA	p-values-CD34 Pro	p-values-CD34-Ery
Differentiation of red blood cells	8.4E-04	1.44E-05
Differentiation of erythroblasts	Not enriched	1.7E-03
Differentiation of leukocytes	2.1E-06	1.5E-03
Development of lymphocytes	4.8E-08	2.5E-03

SMAD1-GREAT	Fold enrichment-CD34-Pro	Fold enrichment-CD34-Ery
Positive regulation of myeloid cell differentiation	3.05	Not enriched
Regulation of lymphocyte differentiation	2.35	Not enriched
Embryonic hematopoiesis	Not enriched	2.84
Erythrocyte differentiation	Not enriched	2.14
Anemia	Not enriched	2.36

E



Trompouki, Bowman et al. Figure S7