The efficiency of hemogenic endothelial specification by RUNX1 is dependent on baseline chromatin accessibility of RUNX1-regulated TGFβ target genes

Elizabeth D. Howell[1], Amanda D. Yzaguirre[1,2], Peng Gao[3,4], Raphael Lis[5,6], Bing He[3,7], Melike Lakadamyali[8], Shahin Rafii[5], Kai Tan[3]*, and Nancy A. Speck[1]*

[1]Abramson Family Cancer Research Institute, Department of Cell and Developmental Biology, Institute for Regenerative Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA 19104, USA

[2]Current address: Fate Therapeutics, 3535 General Atomics Court, San Diego, CA 92121, USA

[3]Departments of Pediatrics, Cell and Developmental Biology and Genetics, Children's Hospital of Philadelphia and University of Pennsylvania, Philadelphia, PA 19104, USA

[4]Current address: The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shanxi 710061, China

[5]Ansary Stem Cell Institute, Department of Genetic Medicine, and Howard Hughes Medical Institute, Weill Cornell Medical College, New York, New York 10065, USA

[6]Current address: Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine, Weill Cornell Medicine, New York, NY 10065, USA

[7]Current address: Department of Pediatrics, College of Medicine, Pennsylvania State University, Hershey, PA 17033, USA

[8]Department of Physiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA 19104, USA

*Corresponding authors
nancyas@upenn.edu
tank1@chop.edu

Supplemental Materials

SUPPLEMENTAL METHODS

Lineage assays. 5,000 c-KIT$^+$ ECs sorted from E9.5 or E13.5 mice were plated per well in a 12 well plate. Sorted ECs were cocultured on a confluent monolayer of OP9 cells with alpha MEM media containing 10% FBS, antibiotics, 10ng/mL rm SCF, rm IL3, rm Flt3L, rm VEGF and rm IL7 to determine B cell, myeloid and erythroid potential. Sorted ECs were cocultured on a confluent monolayer of OP9 stromal cells expressing the Notch ligand delta-like ligand 1 (OP9-DLL1) with alpha MEM media containing 10% FBS, Penicillin-Streptomycin (1X), 10ng/mL rm SCF, rm IL3, rm Flt3L, rm VEGF and rm IL7 to determine T cell potential. T cell cultures were analyzed 9-11 days after plating and B cell cultures were analyzed 10-14 days after plating via flow cytometry. T cells were identified as CD45$^+$ CD25$^+$ CD90$^+$ cells and B cells as CD45$^+$B220$^+$ cells. For hematopoietic progenitor (HP) assays, sorted ECs from E9.5 or E13.5 mice were plated in limiting dilutions in 96-well plates in appropriate cell numbers to reach limiting range as determined empirically in previous limiting dilution assays. Cultures were scored using a light microscope for hematopoietic growth. Wells that fell within the limiting range (less than 30% of total wells for that dilution had hematopoietic growth) were transferred to methylcellulose. Cultures were incubated at 37°C, 5% CO_2. Colonies were scored 7-10 days later.

Endothelial cell HP assay. Freshly sorted cells were plated in methylcellulose (Stem Cell Technologies) in duplicate and incubated at 37°C, 5% CO_2. Colonies were scored 7-10 days later.

Postnatal HEC assay on an instructive vascular niche monolayer. One-month-old postnatal +2Runx1 and Ctrl littermates were fed 2mg of tamoxifen via oral gavage, euthanized 24 hours later, and livers were harvested. ECs were sorted from the livers and plated in multiple dilutions on a vascular niche monolayer as previously described (Sandler et al. 2014). In brief, the vascular

niche monolayer consists of ECs transduced with the adenoviral *E4ORF1* gene which allows the cells to be cultured in the absence of serum. The cells were cultured in hematopoietic media containing StemSpan SFEM, 10% KnockOut serum replacement, 5ng/mL FGF-2 (PeproTech), 10ng/ml EGF (PeproTech), 20ng/mL SCF (PeproTech), 20ng/mL Flt3L (PeproTech), 20ng/mL TPO (PeproTech), 20ng/mL IGF-1 (PeproTech), 10ng/mL IGF-2 (R&D), 10ng/mL IL3 (PeproTech), and 10ng/mL IL-6 (PeproTech) and observed for 20 days for the presence of hematopoietic cells.

RNA-Seq. Cells were sorted into Trizol-LS (Thermo Fisher Scientific) and total RNA was purified using RNeasy micro kit (Qiagen). RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies) and RNA integrity determined with Agilent TapeStation. RNA library preparations, sequencing reactions, and initial analysis were conducted at GENEWIZ, LLC. SMART-Seq v4 Ultra Low Input Kit for Sequencing (Takara) was used for full-length cDNA synthesis and amplification, and Illumina Nextera XT library was used for sequencing library preparation. Briefly, cDNA was fragmented, and an adaptor was added using Transposase, followed by limited-cycle PCR to enrich and add an index to the cDNA fragments. The final library was assessed with Qubit 2.0 Fluorometer and Agilent TapeStation. Libraries were sequenced on Illumina HiSeq2500 using a 2x150 Paired End (PE) configuration. Each sample had two biological replicates. One of the replicates of E13.5 Ctrl ECs deviated significantly from all other embryonic and fetal RNA-seq samples and was discarded. Only one replicate was used for E13.5 Ctrl ECs.

ATAC-seq. ATAC-seq libraries were prepared as described (Buenrostro et al. 2013). In brief, 50,000 ECs were collected and washed with cold PBS. Cell pellets were resuspended in 50 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% (v/v) Igepal CA-630) and immediately centrifuged at 1,600g, 4 °C for 10 min. Nuclei pellets were resuspended in 50 µl of transposition reaction mix (1× Tagment DNA Buffer, 2.5 µl of Tagment DNA Enzyme 1)

and incubated for 30 min at 37 °C. Subsequent steps of the protocol were performed as previously described (Buenrostro et al. 2013). Libraries were purified using a Qiagen MinElute Gel Purification kit and the concentrations were measured using both Qubit and KAPA qPCR. Agilent Bioanalyzer 2100 was used to determine the quality of libraries. Libraries were sequenced on the Illumina NextSeq 500, with 75-bp paired-end reads. Each sample had two biological replicates.

H3K4me1 ChIP-Seq. Purified cells were fixed and crosslinked in 1% formaldehyde in 1× Fixing Buffer for 5 min at room temperature according to the vendor's protocol (Covaris). Cells were then stored at -80^{0}C. Crosslinked cells were thawed on ice (100K cells) and resuspended in 1× Shearing Buffer and sonicated with Covaris E220 for 720s using the following settings: 5% duty factor, 105W Peak Incident Power, and 200 cycles per burst. 10% of sheared chromatin was used as the input and the remaining chromatin was divided into two equal aliquots for immunoprecipitation (50,000 cells per IP). IPs were performed using ChIP-IT high sensitivity kit (Active Motif). IP and input samples were treated with RNase A followed by proteinase K. Cross-linking was reversed by incubation overnight at 65^{0}C and DNA was purified using a MinElute PCR purification kit (Qiagen). All IP DNA and 1-2 ng of input DNA were used for library preparation using the ThruPLEX DNA-Seq kit and Smarter DNA single index kit (Takara). 13 and 9 cycles were used for IP DNA and input DNA, respectively at step 5. Following library amplification, the libraries were purified using bead purification. The concentrations were measured using both Qubit and KAPA qPCR. Agilent Bioanalyzer 2100 was used to check the quality of libraries. Libraries were sequenced on Illumina HiSeq 2500 sequencer in single-end mode with read length of 75bp.

RNA-seq data processing. Sequencing reads were demultiplexed using Bcl2Fastq v2.20 then trimmed and filtered for quality using Trim Galore v0.6.4 (Martin 2011) with the following settings: fastqc, paired, and trim1. Reads were then aligned to the mouse genome (mm10) using STAR

2.7.3a (Dobin et al. 2013) and the mm10 gene annotation file from UCSC. Only uniquely mapped reads with fewer than 2 mismatches were used for downstream analyses. Samtools v1.1 (Danecek et al. 2021) was used to convert to BAM files, and Sambamba v0.6.6 (Tarasov et al. 2015) was used to filter out duplicates, multi-mappers, and unmapped reads. The featureCounts function of the Subread v2.0.0 package (Liao et al. 2014) was used to extract gene-level and transcript level read counts. Fragments per kilobase of transcript per million (FPKM) counts were determined using normalized read counts (TMM and EdgeR) and gene transcript lengths. Differential expression was performed using EdgeR v3.22.5 (Robinson et al. 2009; McCarthy et al. 2012) (FDR < 0.05 and fold change > |1.5|). Bigwig files for visualization were created using Deeptools v3.3.0 (Ramírez et al. 2016) from merged BAM files. The following parameters were used: normalized to reads per genomic content (RPGC), effective genome size: 2,308,125,349 bp, ignore for normalization: ChrX, min fragment length: 20, bin size: 10. Integrated Genome Viewer (IGV) v2.8.9 (Robinson et al. 2011; Thorvaldsdóttir et al. 2013) and pheatmap v1.0.12 in R were used for visualization.

ATAC-seq data processing. Sequencing reads were demultiplexed using Bcl2Fastq v2.20 then trimmed and filtered for quality using Trim Galore v0.6.4 (Martin 2011) with the following settings: fastqc, paired, and trim1. Reads were then aligned to the mouse genome (mm10) using bowtie 2.3.5.1 (Langmead et al. 2019; Langmead and Salzberg 2012). Only uniquely mapped reads with fewer than 2 mismatches were used for downstream analyses. Samtools v.1.1 (Danecek et al. 2021) was used to convert SAM files to BAM files, and Sambamba v0.6.6 (Tarasov et al. 2015) was used to filter out duplicates, multi-mappers, reads mapped to ChrM or blacklist regions, and unmapped reads. MACS2 2.1.4 (Zhang et al. 2008) was used for peak calling using the following parameters: BAMPE, q: 0.05. Differential peaks were determined using DiffBind v2.15.1 and EdgeR (FDR < 0.05) (Stark, R and Brown 2016; Ross-Innes et al. 2012). Merged replicates were used to create bigwig files for visualization using Deeptools v3.3.0 (Ramírez et al. 2016) and the

following parameters: normalized to reads per genomic content (RPGC), effective genome size: 2,308,125,349 bp, ignore for normalization: ChrX, min fragment length: 20, bin size: 10. IGV was used for visualization. Deeptools was also used to plot regions of differential peaks. GREAT 4.0.4 (Hiller et al. 2013; McLean et al. 2010) was used for linking peak regions to genes and subsequent gene ontology annotation using the following parameters: species assembly: mm10; association rule: basal+extension: 5000 bp upstream, 1000 bp downstream, 1,000,000 bp max extension, curated regulatory domains included. Homer v4.11 (Heinz et al. 2010) was used for genomic annotation of peak regions. Motif enrichment of differential accessible peaks was done using MEME Analysis of Motif Enrichment (AME) v5.5.1 (McLeay and Bailey 2010) using shared peaks between samples as the background region.

ATAC-seq footprinting analysis. ATAC-seq footprinting analysis was performed using the Regulatory Genomics Toolbox (RGT) and HMM-based IdeNtification of Transcription factor footprints (HINT) software (Li et al. 2019). In brief, footprints were called from regions of chromatin accessibility (peaks called by MACS2 2.4.1 on merged replicates) for each sample. Called footprints were then matched to TF motifs using the JASPAR 2020 vertebrate motif database. SCL (GSE13511) and FLI1 (GSE20898) motifs were added from Homer database (Heinz et al. 2010). Motif footprints were then filtered to remove any transcription factors that were expressed at < 1.5 FPKM. Differential activity of transcription factors, as well as plots of each transcription factor footprint, were done using HINT-differential. p-value and read counts were calculated using RGT HINT-differential for each transcription factor. Enriched transcription factor footprints at regions of chromatin with increased accessibility and corresponding p-values were determined using BiFET v.10.0 software (Youn et al. 2019). RUNX1 footprints were extracted for more in-depth analysis; Homer v4.11 (Heinz et al. 2010) was used for genomic annotation, bedtools v2.29.2 (Quinlan and Hall 2010) was used for differential peak annotation, genomic location annotation, and footprint overlap. RUNX1 motif scores were determined by RGT motif analysis

(Gusmao et al. 2016). Combination of motifs near RUNX1 footprints was done using PC-TraFF (Meckbach et al. 2015) using RUNX1 footprint locations as the input and the following parameters: upstream: 1000bp, downstream 1000bp, min distance between motifs: 5bp, max distance between motifs: 200bp.

ChIP-seq Data Processing. Sequencing reads were demultiplexed using Bcl2Fastq v2.20 then trimmed and filtered for quality using Trim Galore (Martin 2011) with the following settings: fastqc, paired, and trim1. Reads were then aligned to the mouse genome (mm10) using bowtie 2.3.5.1 (Langmead et al. 2019; Langmead and Salzberg 2012). Only uniquely mapped reads with fewer than 2 mismatches were used for downstream analyses. Samtools v.1.1 (Danecek et al. 2021) was used to convert SAM files to BAM files, and Sambamba v0.6.6 (Tarasov et al. 2015) was used to filter out duplicates, multi-mappers, reads mapped to ChrM or blacklist regions, and unmapped reads. MACS2 2.1.4 (Zhang et al. 2008) was used for peak calling using the following parameters: broad, q: 0.05. Differential peaks were determined using DiffBind v2.15.1 (Ross-Innes et al. 2012) and DESeq2 (FDR < 0.1) (Love et al. 2014). Merged replicates were used to create bigwig files for visualization using Deeptools v3.3.0 (Ramírez et al. 2016) and the following parameters: normalized to reads per genomic content (RPGC), effective genome size: 2,308,125,349 bp, ignore for normalization: ChrX, min fragment length: 20, bin size: 10, smooth length: 60, extend reads: 150. IGV was used for visualization. Deeptools was also used to plot regions of differential peaks and RUNX1 footprint regions. GREAT 4.0.4 (Hiller et al. 2013; McLean et al. 2010) was used for linking peak regions to genes and subsequent gene ontology annotation using the following parameters: species assembly: mm10; association rule: basal+extension: 5000 bp upstream, 1000 bp downstream, 1,000,000 bp max extension, curated regulatory domains included.

STORM. Sorted ECs ($CD41^-CD45^-Ter119^-CD31^+CD144^+$) were plated on poly-d lysine coated LabTek-II chambered 8 well plates at a density of 50K cells/well. Cells were fixed using ethanol-methanol (1:1) fixation for 6 min at -20^0C. Cells were then permeabilized with 0.2% v/v Triton-X-100 in PBS for 10 min at room temperature with rocking. Cells were blocked with 10% (wt/vol) BSA in PBS for 1 hour at room temperature. Cells were stained overnight at 4^o C with H2b primary antibody (1:25) in blocking buffer. Cells were then washed and stained with conjugated secondary (1:1000) antibody for 40 min. Secondary antibodies were labeled with activator-reporter dye pairs (Alexa Fluor 405-Alexa Fluor 647) in-house, as previously described (Bates et al. 2007). Briefly, the dyes were purchased as NHS ester derivatives: Alexa Fluor 405 Carboxylic Acid Succinimidyl Ester (Invitrogen) and Alexa Fluor 647 Carboxylic Acid succinimidyl Ester (Invitrogen). Antibody labeling reactions were performed by incubating for 40 min at room temperature in a mixture containing the secondary antibody, $NaHCO_3$, and the appropriate pair of activator/reporter dyes diluted in DMSO. Purification of labeled antibodies was performed using NAP5 Columns (GE HealthCare). The dye to antibody ratio was quantified using Nanodrop and only antibodies with a composition of 3-4 Alexa Fluor 405 and 0.9-1.2 Alexa Fluor 647 per antibody were used for imaging.

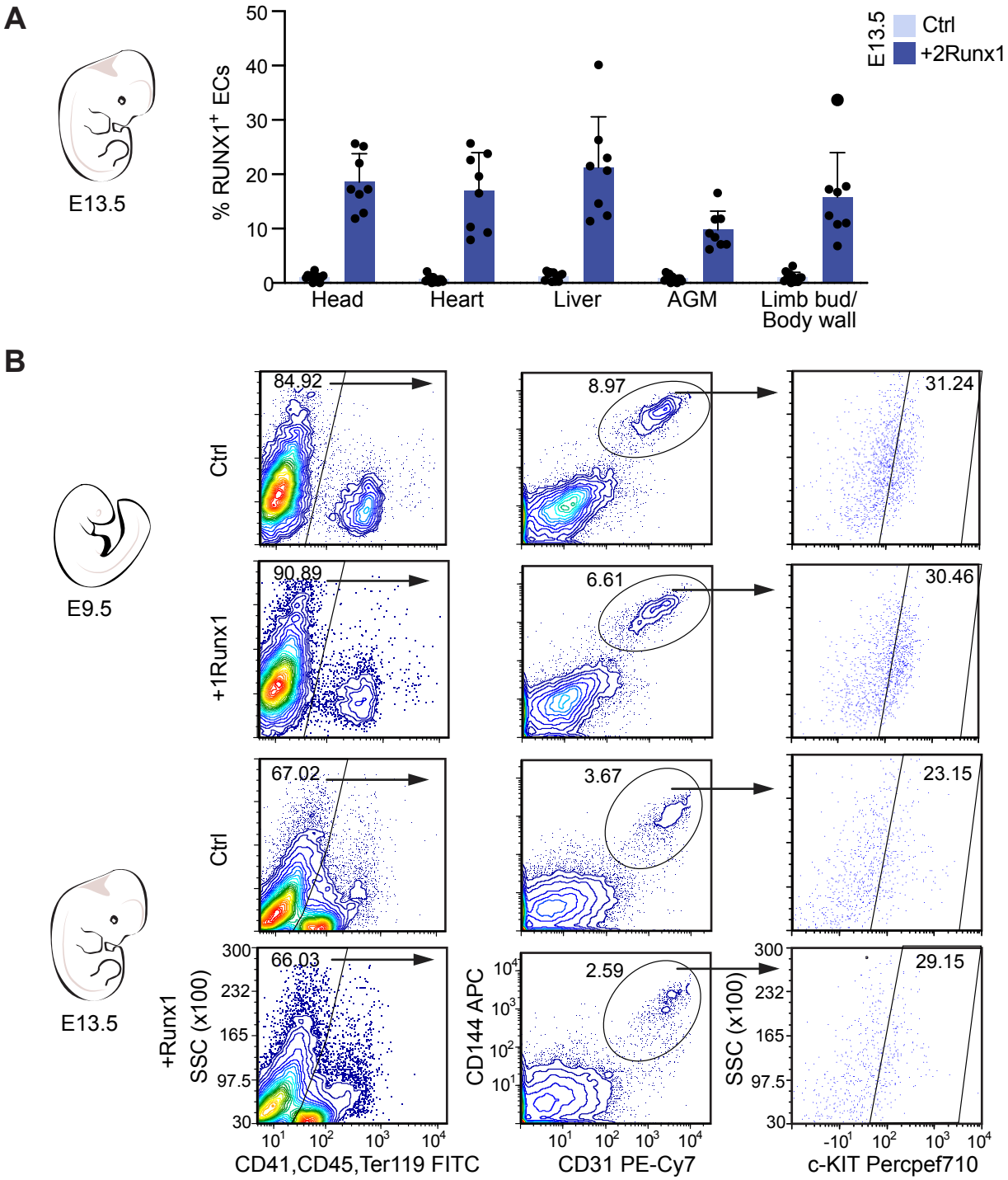
Cells were imaged on an Nanoimager S from Oxford Nanoimaging Limited (ONI) in imaging buffer: cysteamine MEA, glox Solution:0.5 $mgml^{-1}$ glucose oxidase, 40 $mgml^{-1}$ catalase, and 10% Glucose in PBS (Bates et al. 2007). Laser light at 647 nm was used for exciting Alexa Fluor 647 and subsequently switching it off into the dark state, and laser light at 405 nm was used for activating it via an activator dye (Alexa Fluor 405)-facilitated manner. Raw data were analyzed to obtain fluorophore x,y positions using the Nanoimager analysis software. Post-processing of the STORM images was carried out as follows: raw localization data were exported from ONI software, then processed in MATLAB to convert to a bin file, data were then manually cropped using FIJI (Schindelin et al. 2012). Insight3 (provided by Bo Huang, University of California, San Francisco) was used for creating density color coded images. FIJI was used to create signal

intensity plots across the nucleus. A custom written MATLAB script was used for quantifying the percentage of peripheral localization.

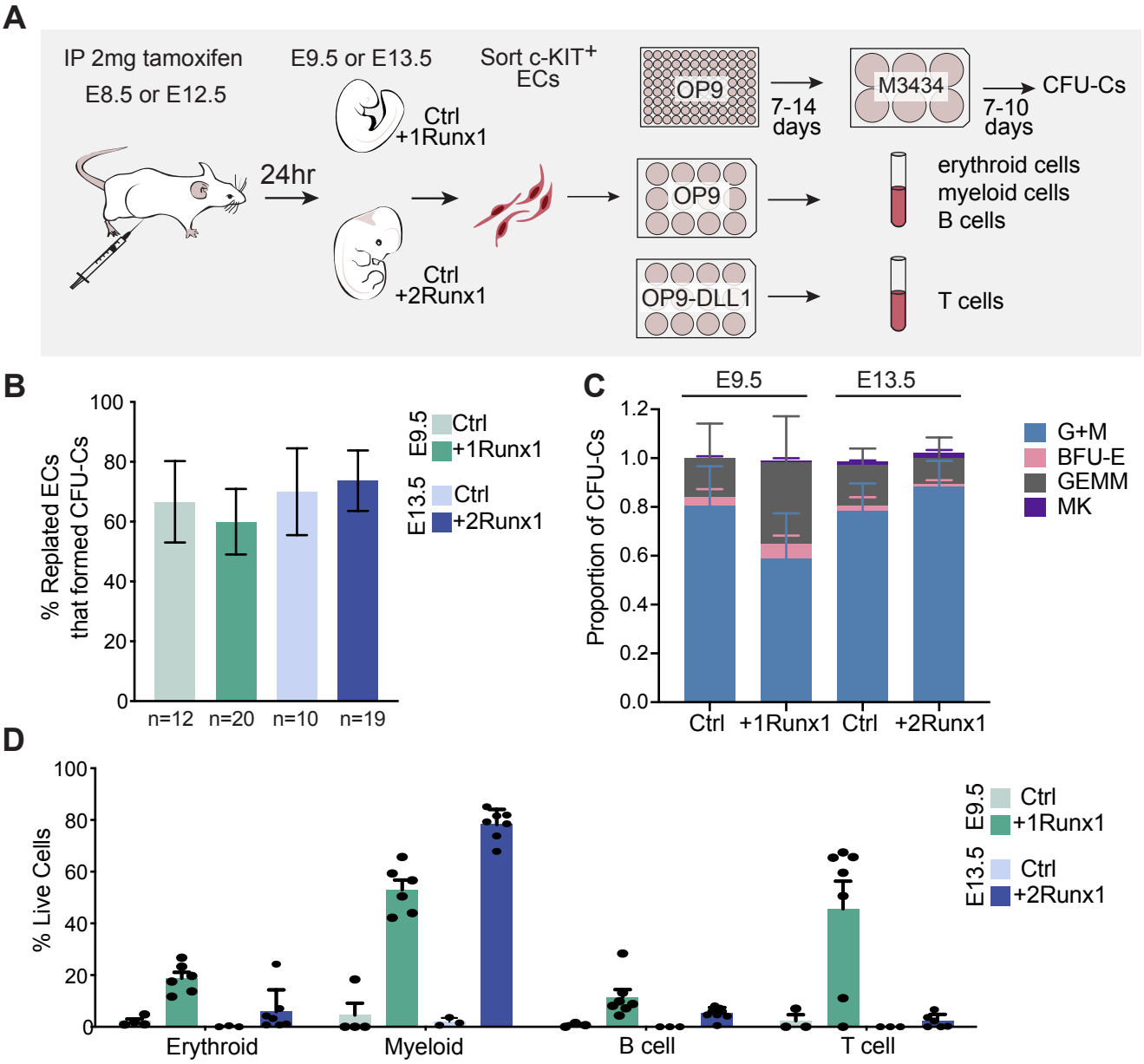
Whole-mount immunofluorescence and confocal microscopy. Embryos were prepared as described previously (Yokomizo et al. 2012). In brief, whole embryos and fetuses were removed from pregnant dams and fixed in 2% paraformaldehyde then dehydrated in methanol. E13.5 fetuses were dissected (head, body wall, and limb buds were removed) prior to fixation. Embryos/fetuses were rehydrated for immunostaining and the following primary antibodies were used: rat anti-mouse CD31 (1:500), rabbit anti-human/mouse RUNX (1:250), rabbit anti-mouse/human phospho-Smad2 (Ser465/467)/Smad3 (Ser423/425) (1:250). Secondary antibodies were goat anti-rabbit Alexa Fluor 647 (1:500), goat-anti rat Alexa Fluor 555 (1:1000) and goat anti-rabbit Alexa Fluor 488 (1:1000). Embryos were dehydrated with methanol then cleared with one part benzyl alcohol and two parts benzyl benzoate. Images were acquired on a Zeiss LSM 710 inverted microscope with ZEN 2011 software. The Zeiss LSM 710 is equipped with 488, 543 and 633 nm wavelengths.

Quantification of pSMAD2/3$^+$ ECs. Images were processed with FIJI software (Schindelin et al. 2012). FIJI Cell Counter application was used to quantify cells. For pSMAD2/3$^+$ ECs, a minimum of 1000 ECs were counted in the endothelial wall then were marked as pSMAD2/3$^+$ or pSMAD2/3$^-$. For pSMAD2/3$^+$ RUNX1$^+$ ECs, a minimum of 350 RUNX1$^+$ ECs were counted in the endothelial wall then were marked as pSMAD2/3$^+$ or pSMAD2/3$^-$.

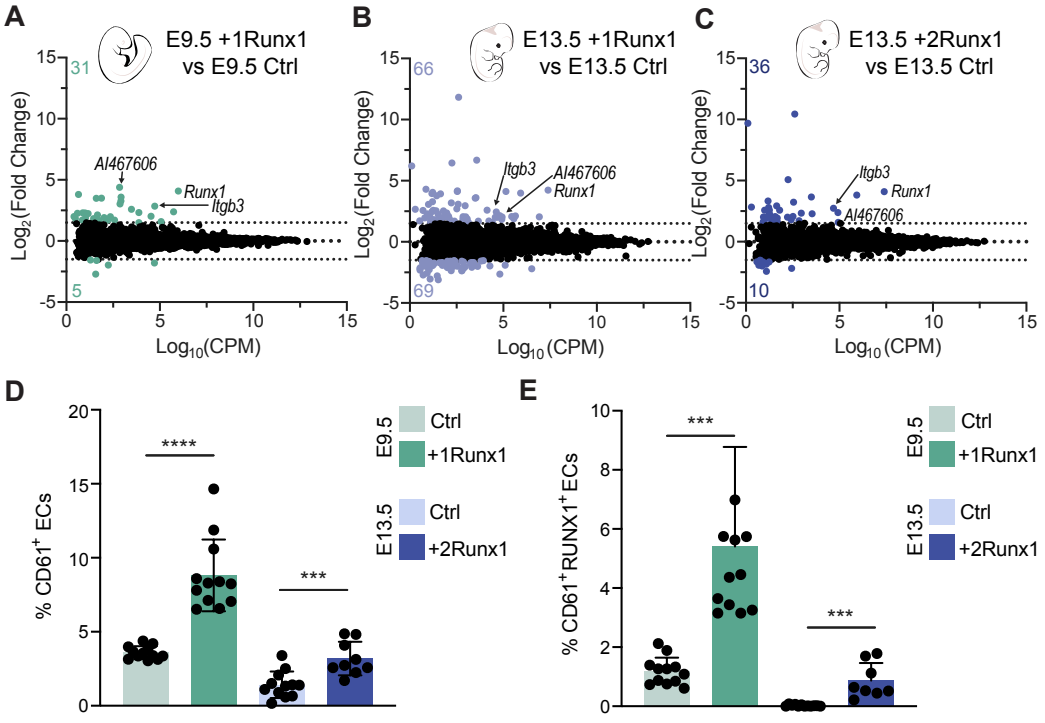
SUPPLEMENTAL FIGURES


A
E13.5
% RUNX1+ ECs
50
40
30
20
10
0
Head
Heart
Liver
AGM
Limb bud/
Body wall
E13.5
Ctrl
+2Runx1
B
E9.5
E13.5
Ctrl
+1Runx1
Ctrl
+Runx1
84.92
8.97
31.24
90.89
6.61
30.46
67.02
3.67
23.15
66.03
2.59
29.15
SSC (x100)
300
232
165
97.5
30
CD41,CD45,Ter119 FITC
CD144 APC
CD31 PE-Cy7
c-KIT Percpef710

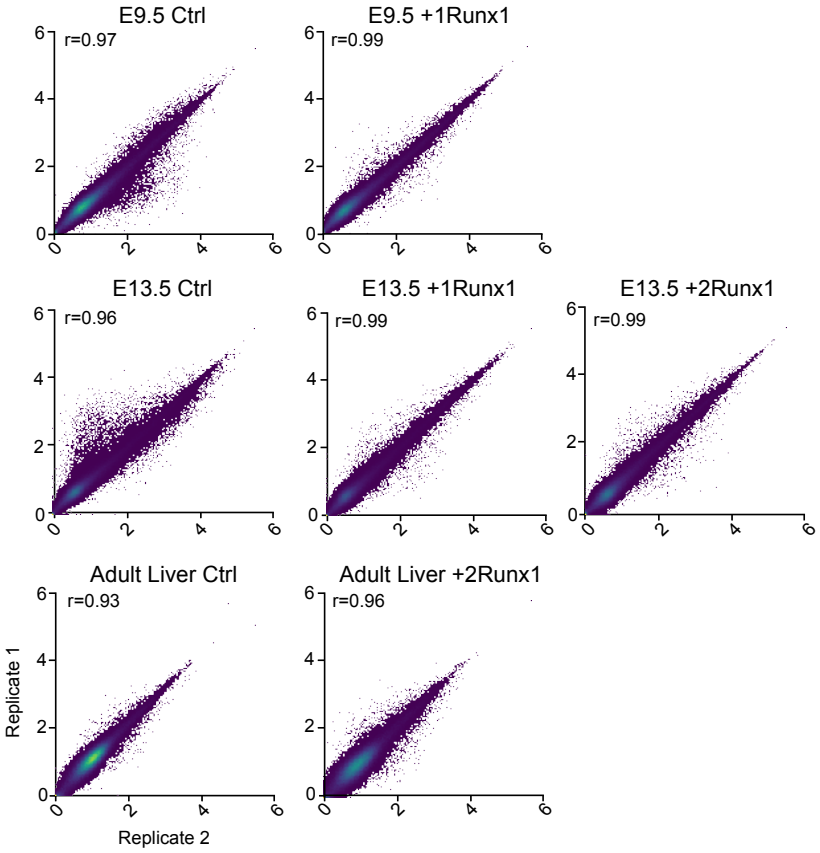
Supplemental Figure S1. Characterization and isolation of embryonic and fetal ECs. A) Bar plot showing percentage (mean ± SD) of RUNX1^{+} ECs (CD41^{-}CD45^{-}Ter119^{-}CD144^{+}CD31^{+}) in individual organs of E13.5 fetuses. Each dot represents a single organ from a single fetus collected from four independent litters/experiments. AGM, aorta-gonad-mesonephros **B)** Representative contour and scatter plots for the isolation of c-KIT^{+} ECs (CD41^{-} CD45^{-} Ter119^{-} CD31^{+}CD144^{+}) from E9.5 embryos (top) and E13.5 fetuses (bottom).


A
IP 2mg tamoxifen
E8.5 or E12.5
24hr
E9.5 or E13.5
Ctrl
+1Runx1
Ctrl
+2Runx1
Sort c-KIT+
ECs
OP9
7-14
days
M3434
7-10
days
CFU-Cs
OP9
erythroid cells
myeloid cells
B cells
OP9-DLL1
T cells
B
% Replated ECs
that formed CFU-Cs
n=12
n=20
n=10
n=19
E9.5
Ctrl
+1Runx1
E13.5
Ctrl
+2Runx1
C
E9.5
E13.5
Proportion of CFU-Cs
Ctrl
+1Runx1
Ctrl
+2Runx1
G+M
BFU-E
GEMM
MK
D
% Live Cells
Erythroid
Myeloid
B cell
T cell
E9.5
Ctrl
+1Runx1
E13.5
Ctrl
+2Runx1

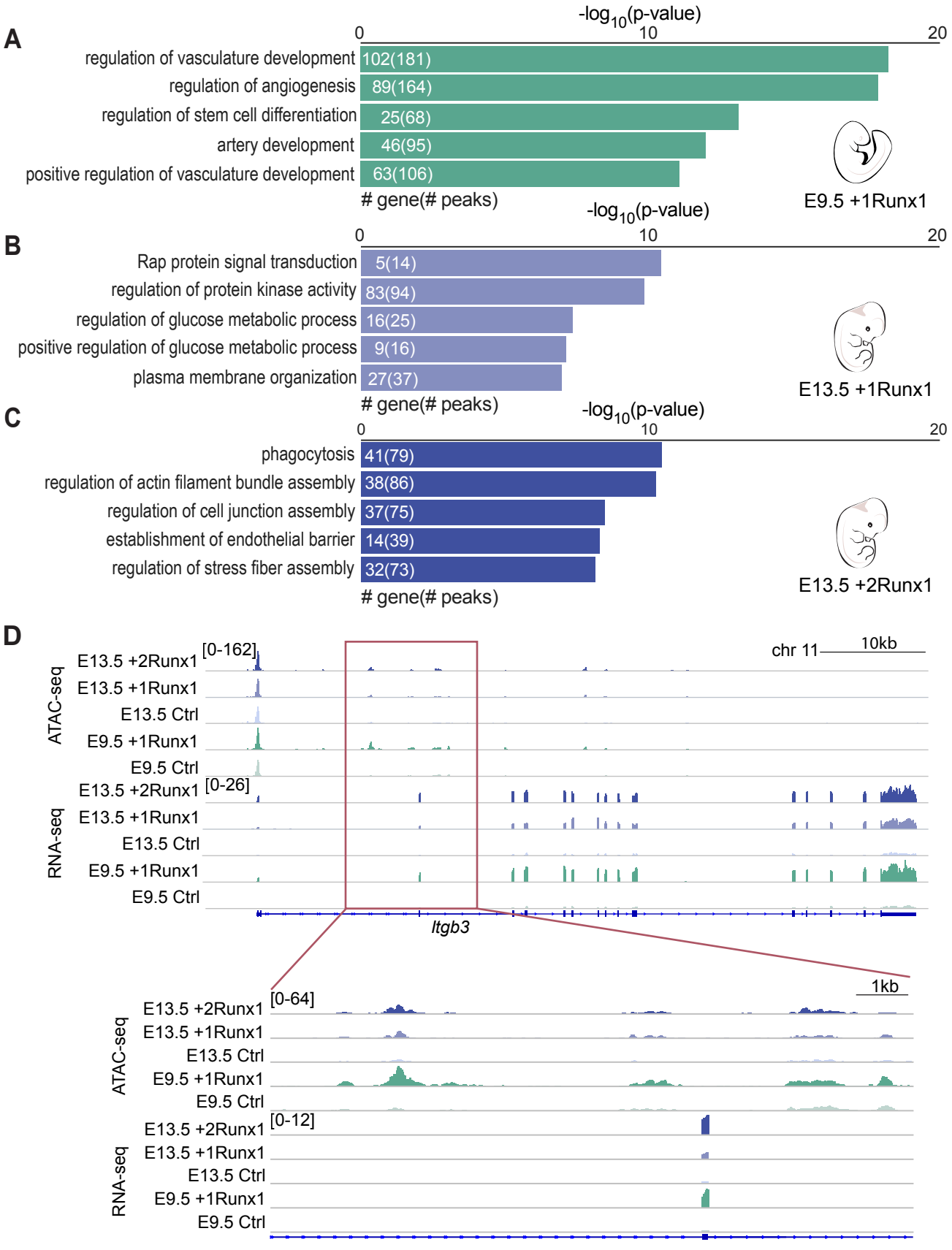
Supplemental Figure S2. RUNX1 expression in ECs yields multilineage hematopoietic progenitors (HPs). A) Schematic illustrating the analysis of embryonic and fetal c-KIT$^+$ ECs (CD41$^-$CD45$^-$Ter119$^-$CD31$^+$CD144$^+$) for their ability to generate colony forming units-culture (CFU-Cs), B cells, erythroid cells, myeloid cells, and T cells. A limiting number of ECs (determined based on HEC frequency in Figure 1E) were plated in each well of a 96-well plate for subsequent CFU-C assays and 5,000 ECs were plated in each well of a 12 well plate containing OP9 or OP9-DLL1 cells for erythroid, myeloid, B and T cell assays. **B)** Percentage (mean ± SD) of wells containing OP9 cells seeded with single ECs that contained HPs capable of forming colonies in methylcellulose cultures. n = number of wells that were assayed for CFU-C activity. **C)** Types of CFU-Cs generated from E9.5 +1Runx1 and E13.5 +2Runx1 ECs. Data are displayed as a proportion of total CFU-Cs (± SD) in each experiment. (MK) megakaryocyte; (GEMM) granulocyte macrophage–erythroid–megakaryocyte; (BFU-E) blast-forming unit-erythroid; (GM) granulocyte macrophage. Data were collected from four and six independent experiments for embryonic and fetal ECs, respectively. **D)** Bar plots (mean ± SD) showing the percentage of phenotypic erythroid (Ter119$^+$), myeloid (Mac1$^+$Gr1$^+$), T (CD45$^+$ CD90$^+$ CD25$^+$) and B (CD45$^+$B220$^+$) cells derived from E9.5 +1Runx1 and E13.5 +2Runx1 ECs. Each point represents a single well derived from 5000 sorted c-KIT$^+$ ECs collected from a minimum of three independent experiments. The reduced T-cell formation from E13.5 +2Runx1 relative to E9.5 +1Runx1 ECs may be due to the lower levels of Notch signaling in E13.5 ECs, as reflected by reduced expression of Notch target genes (*Hey2, Notch3, Gata3*) that are important for T-cell specification (Bellavia et al. 2003; Rowlinson and Gering 2010; Hozumi et al. 2008) in E13.5 ECs.



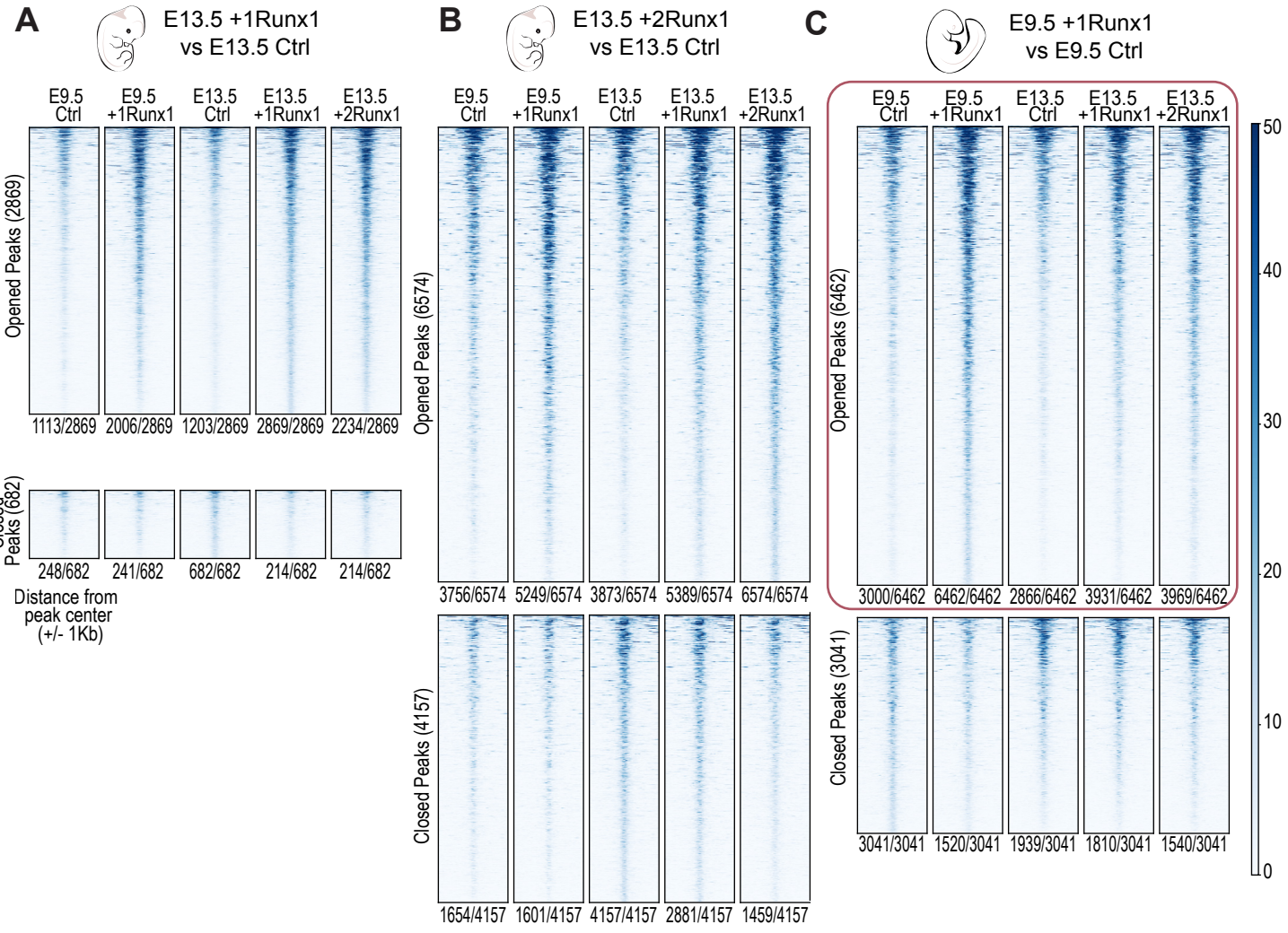
Supplemental Figure S3. RUNX1 induced few transcriptional changes at twenty-four hours post induction. A-C) MA plots following RUNX1 induction showing differentially expressed genes (DEGs) in the upper left of the graphs (relative to Ctrl at each timepoint) for **A**) E9.5 +1Runx1, **B**) E13.5 +1Runx1, and **C**) E13.5 +2Runx1; colored dots and # of genes represent significant DEGs. Selected genes are labeled. **D-E)** Bar plot (mean ± SD) showing percentage of **D)** $CD61^+$ ECs ($CD41^-CD45^-Ter119^-CD144^+CD31^+$) and **E)** $CD61^+RUNX1^+$ ECs. Each point represents pooled embryos (E9.5) or a single fetus (E13.5). Data are from three E9.5 and four E13.5 litters/independent experiments. Unpaired two-tailed Student's t-test was applied to determine significance for both E9.5 and E13.5 samples.

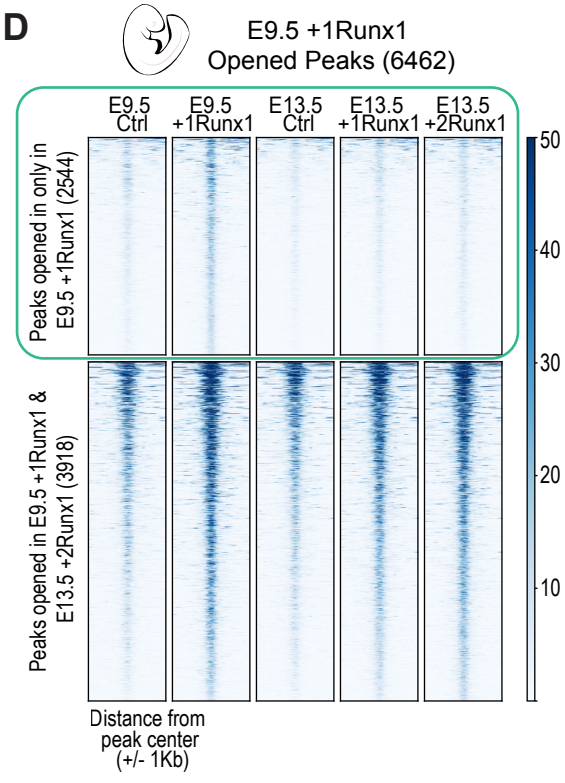


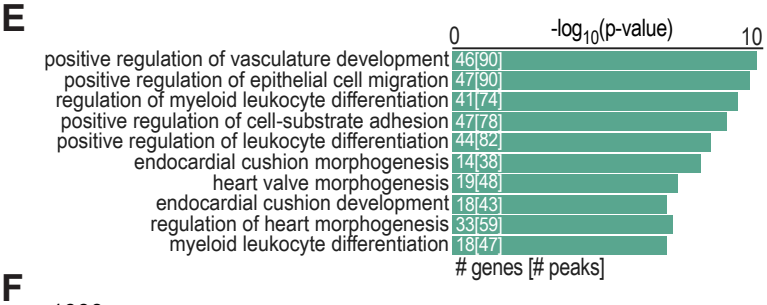
Supplemental Figure S4. Pearson correlation of ATAC-seq replicate samples. Scatter plots of replicates for E9.5 (top), E13.5 (middle), and Adult Liver (bottom) samples in Ctrl and +Runx1 conditions. For each condition, ATAC-seq was performed for two replicates. Pearson correlation values and read scores were calculated using deeptools. The natural log plus one (-log1p) values of the average score for every genomic region (normalized to RPGC, bin size = 10) are plotted for each replicate, and Pearson correlation coefficient values are shown on the upper left of each plot.

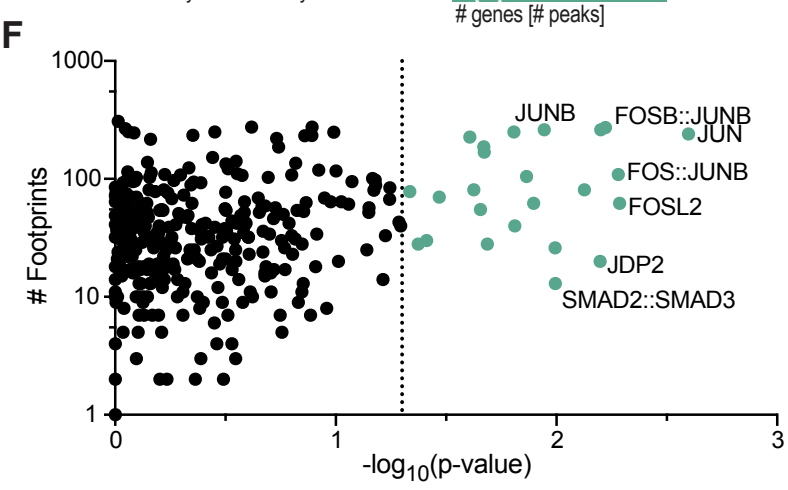

A
-log10(p-value)
0
10
20
regulation of vasculature development
regulation of angiogenesis
regulation of stem cell differentiation
artery development
positive regulation of vasculature development
102(181)
89(164)
25(68)
46(95)
63(106)
# gene(# peaks)
E9.5 +1Runx1
B
-log10(p-value)
0
10
20
Rap protein signal transduction
regulation of protein kinase activity
regulation of glucose metabolic process
positive regulation of glucose metabolic process
plasma membrane organization
5(14)
83(94)
16(25)
9(16)
27(37)
# gene(# peaks)
E13.5 +1Runx1
C
-log10(p-value)
0
10
20
phagocytosis
regulation of actin filament bundle assembly
regulation of cell junction assembly
establishment of endothelial barrier
regulation of stress fiber assembly
41(79)
38(86)
37(75)
14(39)
32(73)
# gene(# peaks)
E13.5 +2Runx1
D
ATAC-seq
RNA-seq
E13.5 +2Runx1
E13.5 +1Runx1
E13.5 Ctrl
E9.5 +1Runx1
E9.5 Ctrl
[0-162]
[0-26]
chr 11
10kb
Itgb3
1kb
[0-64]
[0-12]

Supplemental Figure S5. RUNX1 induction results in changes in chromatin accessibility. **A-C)** GO Biological Processes terms for regions of chromatin with decreased accessibility (relative to Ctrl, FDR<0.05) for **A**) E9.5 +1Runx1, **B**) E13.5 +1Runx1, and **C**) E13.5 +2Runx1 ECs. Top five terms are shown, peak regions were linked to genes using GREAT. Numbers in bars represent number of genes and peaks/regions for each GO term. **D)** Genome browser view of the *Itgb3* gene body (top) and zoomed in view of a specific region that gains chromatin accessibility and increased RNA-seq reads (bottom). ATAC-seq tracks are shown on top and RNA-seq tracks on the bottom. Each sample is shown on the same scale (0-64 for ATAC-seq and 0-12 for RNA-seq) with normalized (to RPGC) tracks.

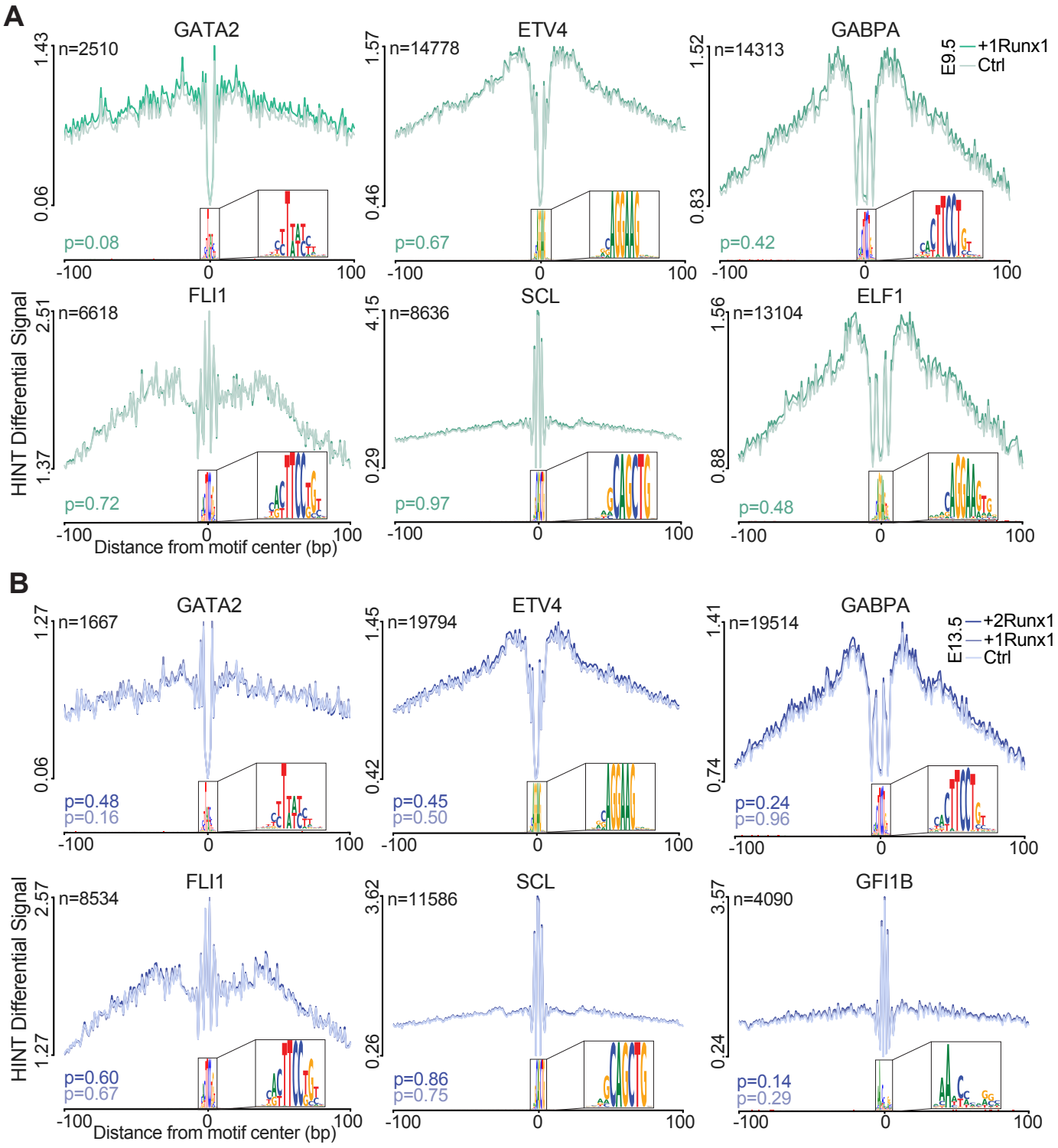

A
E13.5 +1Runx1
vs E13.5 Ctrl
E9.5 Ctrl
E9.5 +1Runx1
E13.5 Ctrl
E13.5 +1Runx1
E13.5 +2Runx1
Opened Peaks (2869)
1113/2869
2006/2869
1203/2869
2869/2869
2234/2869
Closed Peaks (682)
248/682
241/682
682/682
214/682
214/682
Distance from peak center (+/- 1Kb)
B
E13.5 +2Runx1
vs E13.5 Ctrl
Opened Peaks (6574)
3756/6574
5249/6574
3873/6574
5389/6574
6574/6574
Closed Peaks (4157)
1654/4157
1601/4157
4157/4157
2881/4157
1459/4157
C
E9.5 +1Runx1
vs E9.5 Ctrl
Opened Peaks (6462)
3000/6462
6462/6462
2866/6462
3931/6462
3969/6462
Closed Peaks (3041)
3041/3041
1520/3041
1939/3041
1810/3041
1540/3041
50
40
30
20
10
0


D
E9.5 +1Runx1
Opened Peaks (6462)
E9.5 Ctrl
E9.5 +1Runx1
E13.5 Ctrl
E13.5 +1Runx1
E13.5 +2Runx1
Peaks opened in only in E9.5 +1Runx1 (2544)
Peaks opened in E9.5 +1Runx1 & E13.5 +2Runx1 (3918)
50
40
30
20
10
Distance from peak center (+/- 1Kb)

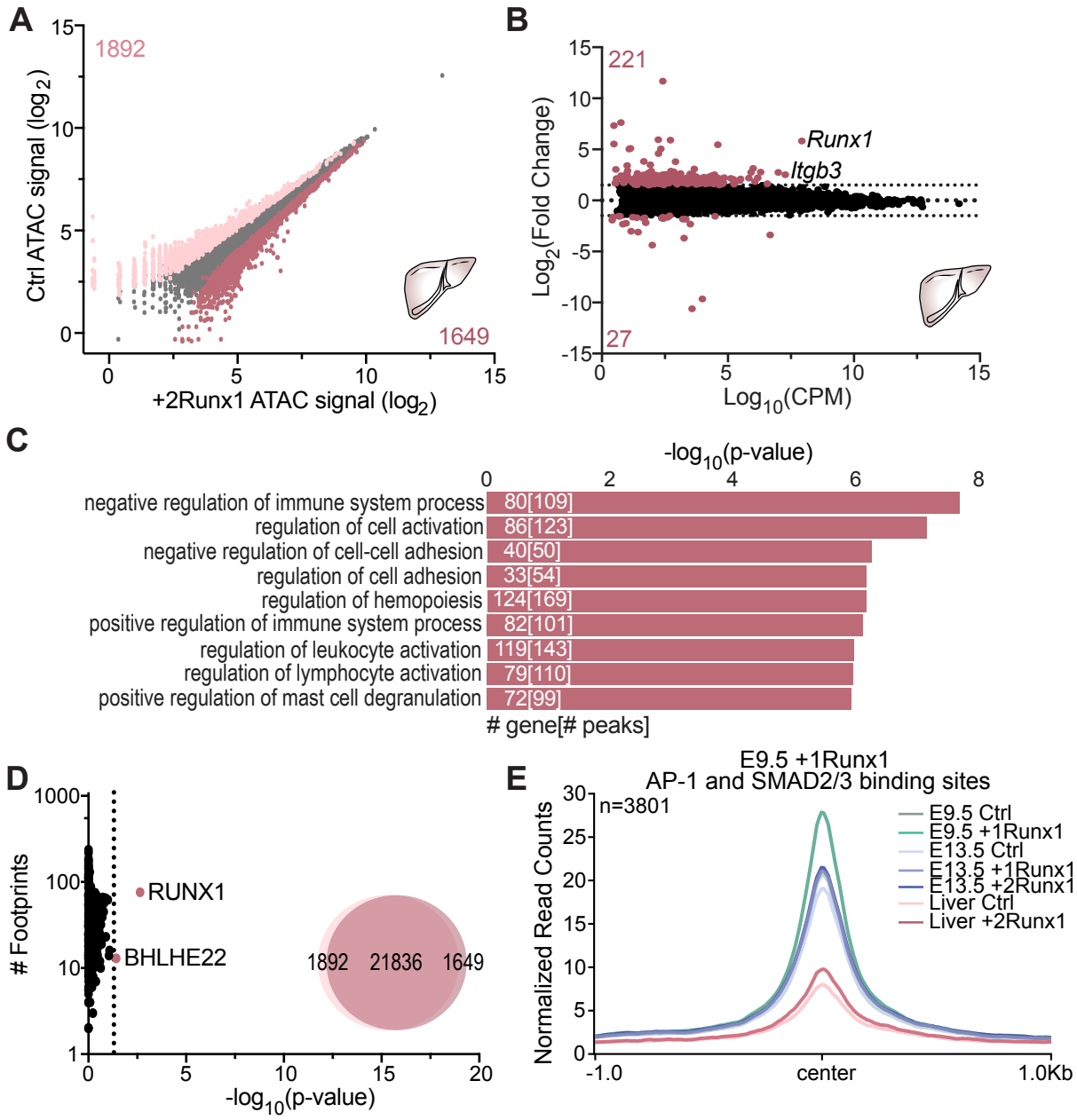

E
-log10(p-value)
0
10
positive regulation of vasculature development 46[90]
positive regulation of epithelial cell migration 47[90]
regulation of myeloid leukocyte differentiation 41[74]
positive regulation of cell-substrate adhesion 47[78]
positive regulation of leukocyte differentiation 44[82]
endocardial cushion morphogenesis 14[38]
heart valve morphogenesis 19[48]
endocardial cushion development 18[43]
regulation of heart morphogenesis 33[59]
myeloid leukocyte differentiation 18[47]
genes [# peaks]


F
1000
100
10
1
Footprints
0
1
2
3
-log10(p-value)
JUNB
FOSB::JUNB
JUN
FOS::JUNB
FOSL2
JDP2
SMAD2::SMAD3

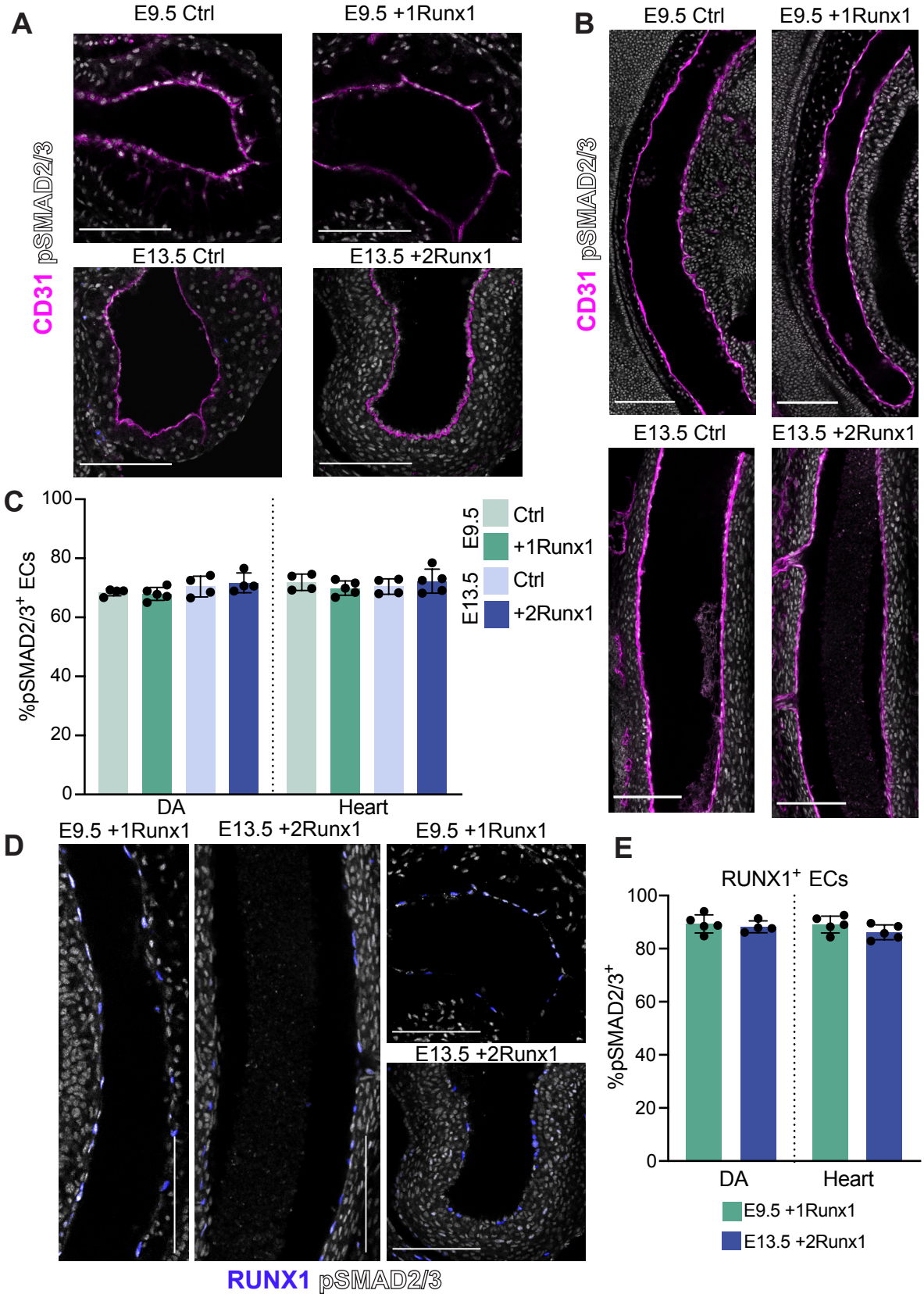
Supplemental Figure S6. RUNX1 induction results in changes in chromatin accessibility. **A-C**) Heatmaps of regions with increased (Opened Peaks, top) or decreased (Closed Peaks, bottom) chromatin accessibility with respect to Ctrl for **A**) E13.5 +1Runx1, **B**) E13.5 +2Runx1, and **C**) E9.5 +1Runx1. Opened and closed peaks were determined using EdgeR (FDR < 0.05) and the number of opened or closed regions is represented as numbers in parentheses on the y-axis. Numbers under the heatmap represent the fraction of open or closed peaks in each sample. Heatmaps were made using deeptools and scale is normalized read count (normalized to RPGC). **D)** Heatmap of regions that gained chromatin accessibility following RUNX1 induction in E9.5 +1Runx1, relative to E9.5 Ctrl ECs, highlighted with a pink box in C). Regions were subsetted into peaks opened only in E9.5 +1Runx1 ECs (top, green box) or peaks opened in both E9.5 +1Runx1 and E13.5 +2Runx1 ECs (bottom). **E)** GO Terms for regions only opened in E9.5 +1Runx1 ECs, highlighted by green box in D). Top 10 terms are shown, and the numbers of genes and regions (in brackets) corresponding to each term are overlaid on the bars. Regions were linked to genes using GREAT. **F)** Scatter plots showing enriched TF footprints at regions only opened in E9.5 +1Runx1 ECs (highlighted in the green box in D), relative to regions opened in both E9.5 +1Runx1 and E13.5 +2Runx1 ECs. The number of footprints for each TF at regions opened only in E9.5 +1Runx1 ECs is displayed on the y-axis. Colored circles indicate $p<0.05$; p-value calculated using biFET.


A
GATA2
n=2510
p=0.08
ETV4
n=14778
p=0.67
GABPA
n=14313
p=0.42
E9.5
+1Runx1
Ctrl
FLI1
n=6618
p=0.72
SCL
n=8636
p=0.97
ELF1
n=13104
p=0.48
HINT Differential Signal
Distance from motif center (bp)
B
GATA2
n=1667
p=0.48
p=0.16
ETV4
n=19794
p=0.45
p=0.50
GABPA
n=19514
p=0.24
p=0.96
E13.5
+2Runx1
+1Runx1
Ctrl
FLI1
n=8534
p=0.60
p=0.67
SCL
n=11586
p=0.86
p=0.75
GFI1B
n=4090
p=0.14
p=0.29
HINT Differential Signal
Distance from motif center (bp)

Supplemental Figure S7. Most TFs enriched at regions of chromatin with increased accessibility do not appear to be significantly altering local chromatin accessibility. A-B) Footprint plots following RUNX1 induction for TFs enriched at regions of chromatin with increased accessibility **A)** GATA2, ETV4, GABPA, FLI1, SCL, and ELF1 in E9.5 ECs and **B)** GATA2, ETV4, GABPA, FLI1, SCL, and GFI1B in E13.5 ECs. Plots show average normalized read counts calculated using HINT-differential around all footprints, n = # of nonredundant footprints, p-value calculated using HINT-differential. The p-value in dark blue corresponds to E13.5 +2Runx1 ECs and light blue corresponds to E13 +1Runx1 ECs, relative to E13.5 Ctrl ECs.


A
1892
1649
Ctrl ATAC signal (log2)
+2Runx1 ATAC signal (log2)
B
221
27
Runx1
Itgb3
Log2(Fold Change)
Log10(CPM)
C
-log10(p-value)
negative regulation of immune system process 80[109]
regulation of cell activation 86[123]
negative regulation of cell-cell adhesion 40[50]
regulation of cell adhesion 33[54]
regulation of hemopoiesis 124[169]
positive regulation of immune system process 82[101]
regulation of leukocyte activation 119[143]
regulation of lymphocyte activation 79[110]
positive regulation of mast cell degranulation 72[99]
gene[# peaks]
D
Footprints
RUNX1
BHLHE22
1892 21836 1649
-log10(p-value)
E
E9.5 +1Runx1
AP-1 and SMAD2/3 binding sites
n=3801
E9.5 Ctrl
E9.5 +1Runx1
E13.5 Ctrl
E13.5 +1Runx1
E13.5 +2Runx1
Liver Ctrl
Liver +2Runx1
Normalized Read Counts
-1.0
center
1.0Kb

Supplemental Figure S8. RUNX1 alone is unable to specify postnatal ECs as hemogenic. **A)** Scatter plot of ATAC-seq peak signals for a consensus peak set in all samples following RUNX1 induction in 1-month postnatal Liver +2Runx1 ECs. Coloring and peak number are based on differential peaks (FDR < 0.05). **B)** MA plots following RUNX1 induction showing differentially expressed genes (DEGs) (relative to control at each timepoint) for Liver +2Runx1 ECs; colored dots and number of genes represent significantly upregulated genes based on fold change > |1.5| as determined by EdgeR. Selected genes are labeled. **C)** GO Biological Processes terms for regions of chromatin with increased accessibility (relative to Ctrl) for 1-month postnatal Liver +2Runx1 ECs. Top ten terms shown, peak regions were linked to genes using GREAT. Numbers in bars represent number of genes and [peaks/regions] for each GO term. **D)** Scatter plots showing enriched TF footprints at regions of chromatin with increased accessibility in 1-month postnatal Liver +2Runx1 relative to Ctrl ECs. The number of footprints for each TF is displayed on the y-axis. p-values were calculated using biFET, colored points indicate p<0.05. Venn diagrams represent the number of peaks closed (left), unchanged (middle), and opened (right) following RUNX1 induction. **E)** Average peak profile plots of normalized (RPGC) ATAC-seq signal for E9.5 +1Runx1 nonredundant AP-1 (FOS::JUNB) and SMAD2/3 footprints for 1 month postnatal Liver +2Runx1, 1 month postnatal Liver Ctrl, E13.5 +2Runx1, E13.5 +1Runx1, E13.5 Ctrl, E9.5 +1Runx1 and E9.5 Ctrl ECs, n=number of footprints.


A
E9.5 Ctrl
E9.5 +1Runx1
E13.5 Ctrl
E13.5 +2Runx1
CD31 pSMAD2/3
B
E9.5 Ctrl
E9.5 +1Runx1
E13.5 Ctrl
E13.5 +2Runx1
CD31 pSMAD2/3
C
%pSMAD2/3+ ECs
100
80
60
40
20
0
DA
Heart
E9.5
Ctrl
+1Runx1
E13.5
Ctrl
+2Runx1
D
E9.5 +1Runx1
E13.5 +2Runx1
E9.5 +1Runx1
E13.5 +2Runx1
RUNX1 pSMAD2/3
E
RUNX1+ ECs
%pSMAD2/3+
100
80
60
40
20
0
DA
Heart
E9.5 +1Runx1
E13.5 +2Runx1

Supplemental Figure S9. The percentage of pSMAD2/3$^+$ ECs is equivalent at E9.5 and E13.5. A-B) Representative confocal cross-sections of **A)** heart ventricle and **B)** dorsal aorta (DA). CD31 is shown in magenta and pSMAD2/3 in white. Scale bars = 100μm. **C)** Quantification (mean $\pm$ SD) of pSMAD2/3$^+$ ECs in the DA (left) and heart (right). **D)** Representative confocal cross-section of DA (left and middle images) and heart (right images) of E9.5 +1Runx1 embryos and E13.5 +2Runx1 fetuses. RUNX1 is shown in blue, and pSMAD2/3 in white. Scale bars = 100μm. **E)** Quantification (mean $\pm$ SD) of pSMAD2/3$^+$ RUNX1$^+$ECs. For C) & E) each dot represents a single E9.5 embryo collected from two independent litters or a single E13.5 fetus collected from two independent litters and four total independent experiments.

SUPPLEMENTAL TABLES

Supplemental Table S1. Frequency of hematopoietic progenitors (HPs) and hemogenic endothelial cells (HECs) in sorted endothelial cell populations.

Stage	Genotype	HP frequency	HEC frequency
E9.5	Ctrl	1:21,803	1:996
	+1Runx1	1:43,600	1:58
E13.5	Ctrl	1:4988	1:9114
	+1Runx1	1:5700	1:1517
	+2Runx1	1:4500	1:185

Supplemental Table S2. RUNX1 "resistant" binding sites and corresponding genes and GO terms for E13.5 fetal ECs that are present in E9.5 embryonic ECs.

Supplemental Table S3. Pairwise tests for differences in HEC frequencies following RUNX1 induction and TGFβ3 treatment

Supplemental Table S4. Antibodies used in this study.

SUPPLEMENTAL REFERENCES

Bates M, Huang B, Dempsey GT, Zhuang X. 2007. Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes. *Science (80-)* **317**: 1749–1753. https://www.sciencemag.org/lookup/doi/10.1126/science.1146598.

Bellavia D, Campese AF, Vacca A, Gulino A, Screpanti I. 2003. Notch3, another Notch in T cell development. *Semin Immunol* **15**: 107–112.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**: 1–4.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Gusmao EG, Allhoff M, Zenke M, Costa IG. 2016. Analysis of computational footprinting methods for DNase sequencing experiments. *Nat Methods* **13**: 303–309.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576–589. http://dx.doi.org/10.1016/j.molcel.2010.05.004.

Hiller M, Agarwal S, Notwell JH, Parikh R, Guturu H, Wenger AM, Bejerano G. 2013. Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: Application to zebrafish. *Nucleic Acids Res* **41**.

Hozumi K, Negishi N, Tsuchiya I, Abe N, Hirano KI, Suzuki D, Yamamoto M, Engel JD, Habu S. 2008. Notch signaling is necessary for GATA3 function in the initiation of T cell development. *Eur J Immunol* **38**: 977–985.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Langmead B, Wilks C, Antonescu V, Charles R. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**: 421–432.

Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. 2019. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* **20**: 45. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1642-2.

Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: An efficient general purpose program for

assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.
Lichtinger M, Ingram R, Hannah R, Müller D, Clarke D, Assi SA, Lie-A-Ling M, Noailles L, Vijayabaskar MS, Wu M, et al. 2012. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J* **31**: 4318–4333.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 1–21.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10. http://journal.embnet.org/index.php/embnetjournal/article/view/200.
McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**: 4288–4297.
McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495–501.
McLeay RC, Bailey TL. 2010. Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**.
Meckbach C, Tacke R, Hua X, Waack S, Wingender E, Gültas M. 2015. PC-TraFF: Identification of potentially collaborating transcription factors using pointwise mutual information. *BMC Bioinformatics* **16**: 1–21. http://dx.doi.org/10.1186/s12859-015-0827-2.
Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160–W165.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**: 389–393.
Rowlinson JM, Gering M. 2010. Hey2 acts upstream of Notch in hematopoietic stem cell specification in zebrafish embryos. *Blood* **116**: 2046–2056.
Sandler VM, Lis R, Liu Y, Kedem A, James D, Elemento O, Butler JM, Scandura JM, Rafii S. 2014. Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature* **511**: 312–318.
Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676. https://doi.org/10.1038/nmeth.2019.
Stark, R and Brown G. 2016. DiffBind: differential binding analysis of ChIP-Seq peak datatle. *http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf* 1–29. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.
Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**: 2032–2034.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.
Yokomizo T, Yamada-Inagawa T, Yzaguirre AD, Chen MJ, Speck NA, Dzierzak E. 2012. Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nat Protoc* **7**: 421–423.
Youn A, Marquez EJ, Lawlor N, Stitzel ML, Ucar D. 2019. BiFET: Sequencing Bias-free transcription factor Footprint Enrichment Test. *Nucleic Acids Res* **47**: 1–14.
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM,

Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**.