Supplemental Figures

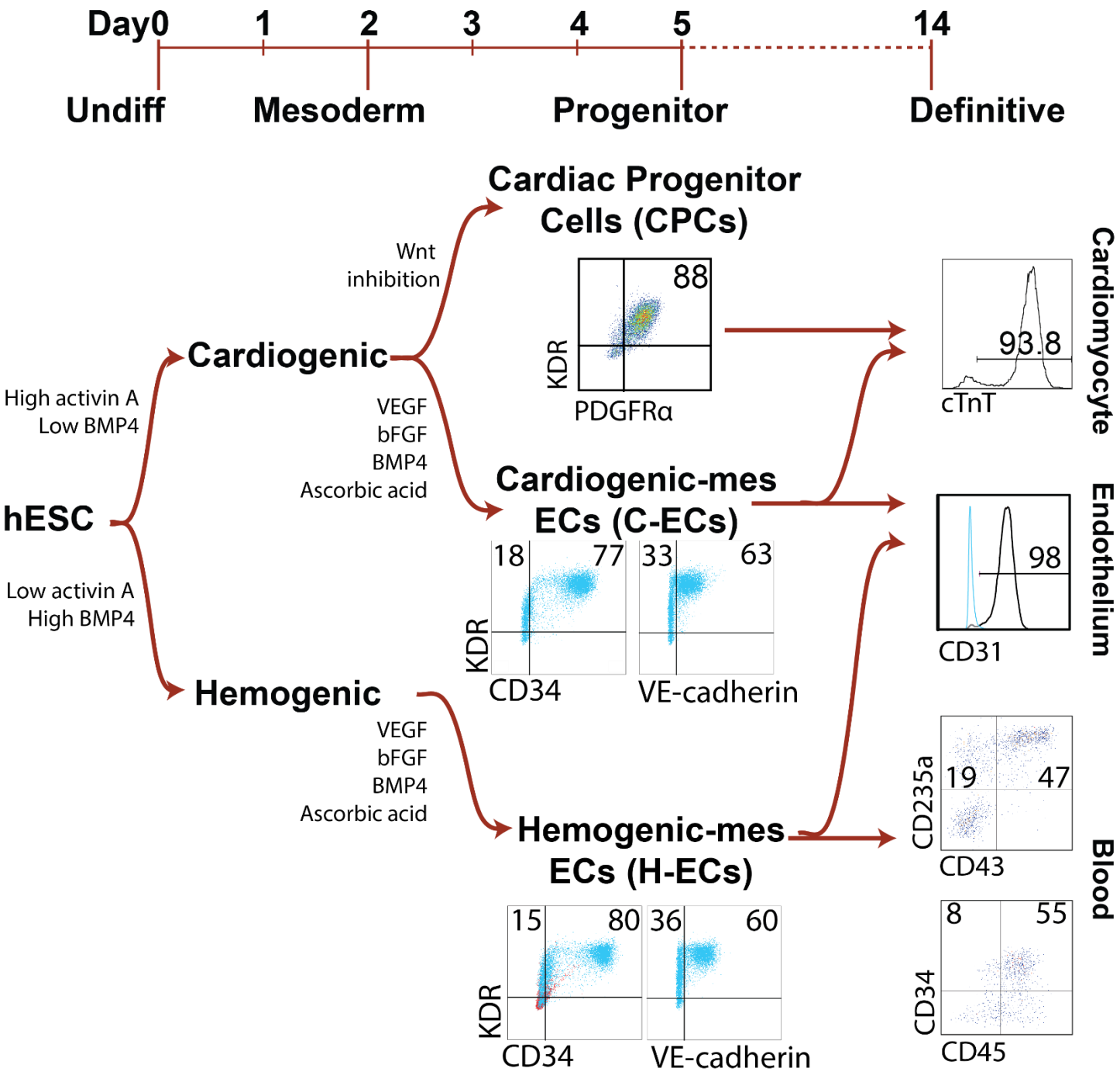

Figure S1 Related to Figure 1. Schematic diagram of differentiation protocols used to generate cardiac and EC lineages. Representative flow plots are shown for enrichment of lineages based on markers including KDR/PDGFRα for the cardiac lineage and KDR/CD34/VE-cadherin for the endothelial lineages. Flow plots of mature cells types are shown including cardiomyocytes based on expression of cTnT, endothelial cells based on expression of CD31, and blood derivatives based on expression of CD235a/CD43 or CD45/CD34.

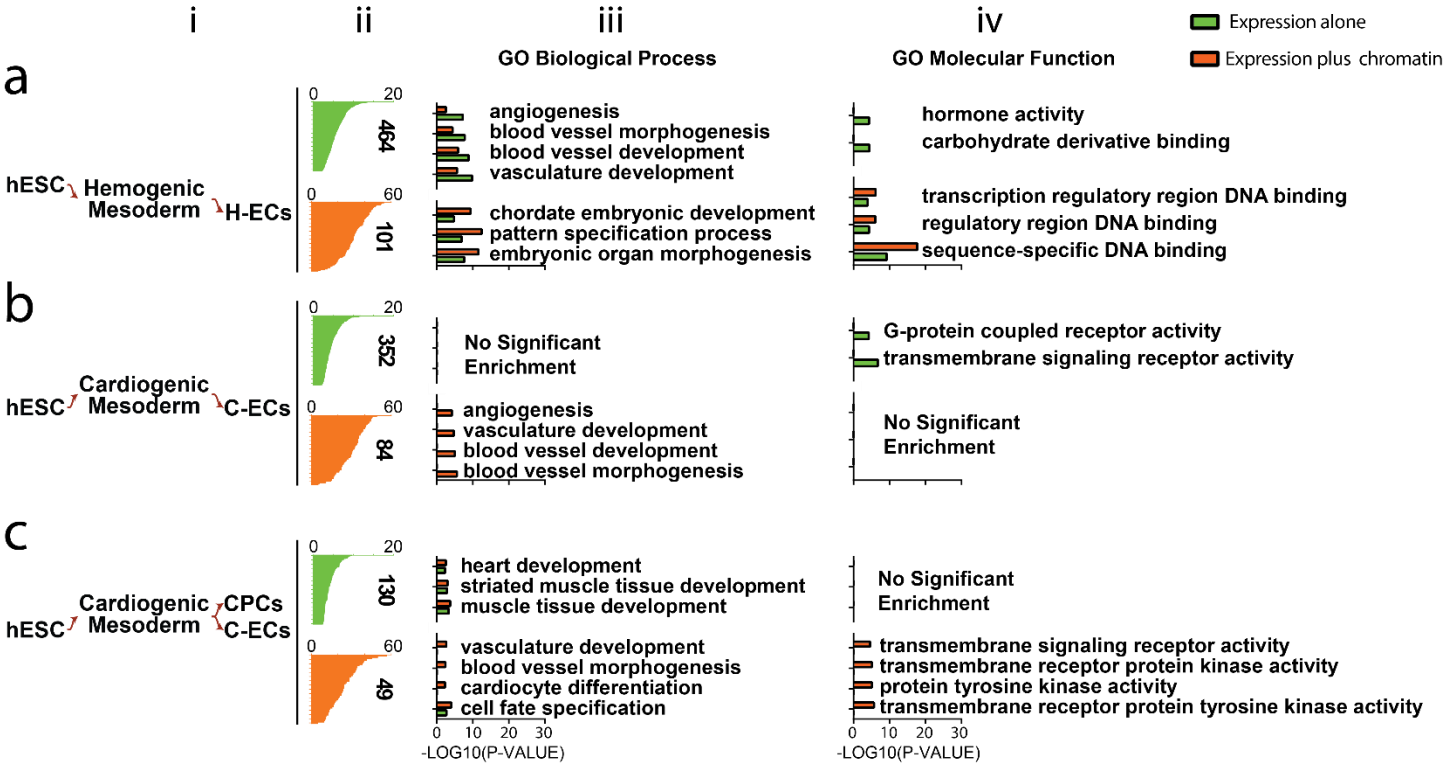

Figure S2 Related to Figure 2. Gene ontology analysis of regulator lists. Regulator lists generated using ranking algorithm outlined in Figure 2a-d were analyzed by gene ontology analysis with data shown for H-ECs (a) and C-ECs (b) and CPCs/C-ECs (c). Data are presented as (i) lineage map of population under evaluation, (ii) ranked analysis of regulators based on expression (green) or expression plus chromatin (orange). The number of genes in each regulator list is shown to the side of each graph. (iii-iv) Gene ontology enrichment categories for Biological Process (iii) and Molecular Function (iv) as a comparison of regulators identified by expression alone (green) or expression plus chromatin (orange).

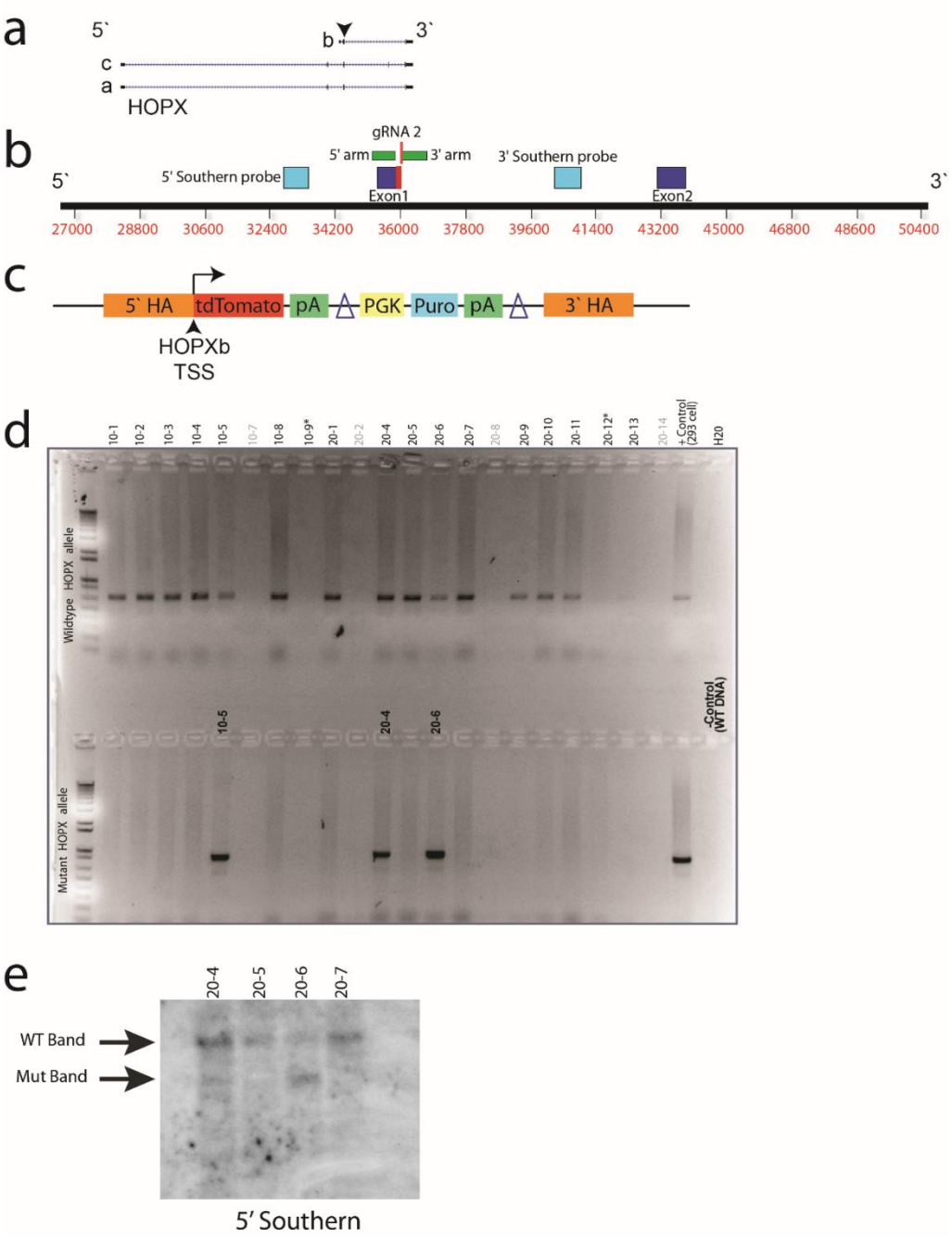

Figure S3 Related to Figure 4. Genotype analysis of HOPX gene targeting in hESCs. (a) Schematic of the HOPX locus with arrow pointing the TSS. (b) HOPX locus with details of gRNA binding site and southern probes. (c) Schematic of targeting vector used for knockin. (d) PCR screening to identify correctly targeted clones. (e) Southern blot results for 4 clones showing the WT band and mutant bands.

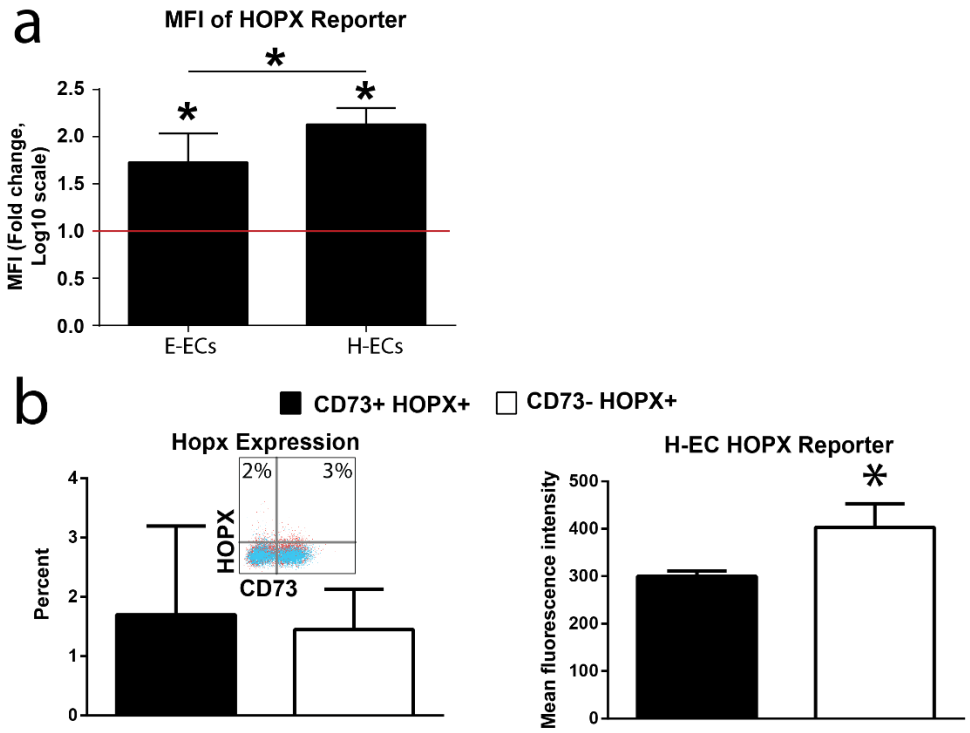

Figure S4 related to Figure 4. HOPX reporter activity in EC differentiation. (a) HOPX-tdTomato reporter activity in day 5 CD34+ C-ECs and H-ECs. HOPX is expressed higher than WT cells with significantly higher reporter activity in H-ECs than C-ECs. (b) Analysis of CD73 cells derived from H-EC differentiation showing no difference in the number of HOPX expressing cells in the CD73+ or CD73- populations (left) but significantly higher HOPX-tdTomato reporter activity in CD73+ vs CD73- cells at day 5.

Genes up-regulated in WTFLK+SCL+ vs SCLKO

Term	Count	PValue	FDR
GO:0005856~cytoskeleton	63	2.17E-07	2.95E-04
alternative splicing	184	6.40E-07	8.77E-04
GO:0007049~cell cycle	40	6.42E-06	1.11E-02
GO:0030097~hemopoiesis	23	6.52E-06	1.12E-02
IPR011993:Pleckstrin homology-type	24	9.71E-06	1.51E-02
GO:0043232~intracellular non-membrane-bounded organelle	86	1.14E-05	1.55E-02
GO:0043228~non-membrane-bounded organelle	86	1.14E-05	1.55E-02
GO:0002520~immune system development	24	2.69E-05	4.64E-02
GO:0030099~myeloid cell differentiation	13	2.79E-05	4.80E-02

Genes down-regulated in WTFLK+SCL+ vs SCLKO

Term	Count	PValue	FDR
GO:0007507~heart development	28	6.33E-11	1.09E-07
GO:0000904~cell morphogenesis involved in differentiation	24	1.45E-08	2.49E-05
GO:0048729~tissue morphogenesis	25	2.93E-08	5.03E-05
GO:0003700~transcription factor activity	47	3.46E-08	5.00E-05
GO:0007498~mesoderm development	13	4.49E-08	7.73E-05
GO:0048738~cardiac muscle tissue development	11	2.55E-06	0.00438
GO:0003007~heart morphogenesis	12	4.37E-06	0.00752
GO:0048634~regulation of muscle development	9	1.66E-05	0.02863
GO:0043565~sequence-specific DNA binding	32	2.36E-05	0.03408

Gene ID	Probe ID	Fold change WTFlk+Scl+ vs SclKO	P-value	Fold change WTFlk+Scl+ vs WTFlk+Scl-	P-value
Cbfa2t3	1438705_at	123.2	3.6E-04	11.0	7.98E-03
Myb	1450194_a_at	96.8	3.7E-04	9.4	5.82E-03
Lyl1	1419120_at	33.0	1.4E-04	6.8	2.37E-03
Sox7	1416564_at	31.9	3.3E-03	4.2	3.57E-02
Gata1	1449232_at	14.9	8.7E-04	6.8	6.36E-03
Runx1	1422864_at	5.4	1.1E-03	2.5	1.71E-02
Gata2	1450333_a_at	5.3	2.1E-03	4.3	1.19E-02
Hopx	1451776_s_at	3.8	9.8E-03	3.4	1.52E-02
Eomes	1426001_at	-10.1	6.8E-04	-7.2	1.59E-03
Isl1	1450723_at	-5.4	2.1E-03	-7.4	1.29E-03
Pdgfra	1421916_at	-4.8	5.7E-04	-2.8	5.86E-03
Mesp1	1426557_at	-4.2	5.1E-04	-4.5	4.58E-04
Gata4	1418864_at	-4.2	1.8E-03	-2.7	1.46E-02
Dkk1	1420360_at	-4.2	2.3E-04	-2.7	1.65E-03

Figure S5 related to Figure 7. HOPX is up-regulated in an SCL-dependent manner. Data derived from Org et al showing gene ontologies up-regulated and down-regulated in WTFLK+SCL+ vs SCL knockout cells (left) and selected genes reflecting these gene ontologies (right) highlighting HOPX significantly up-regulated only in SCL replete cells.

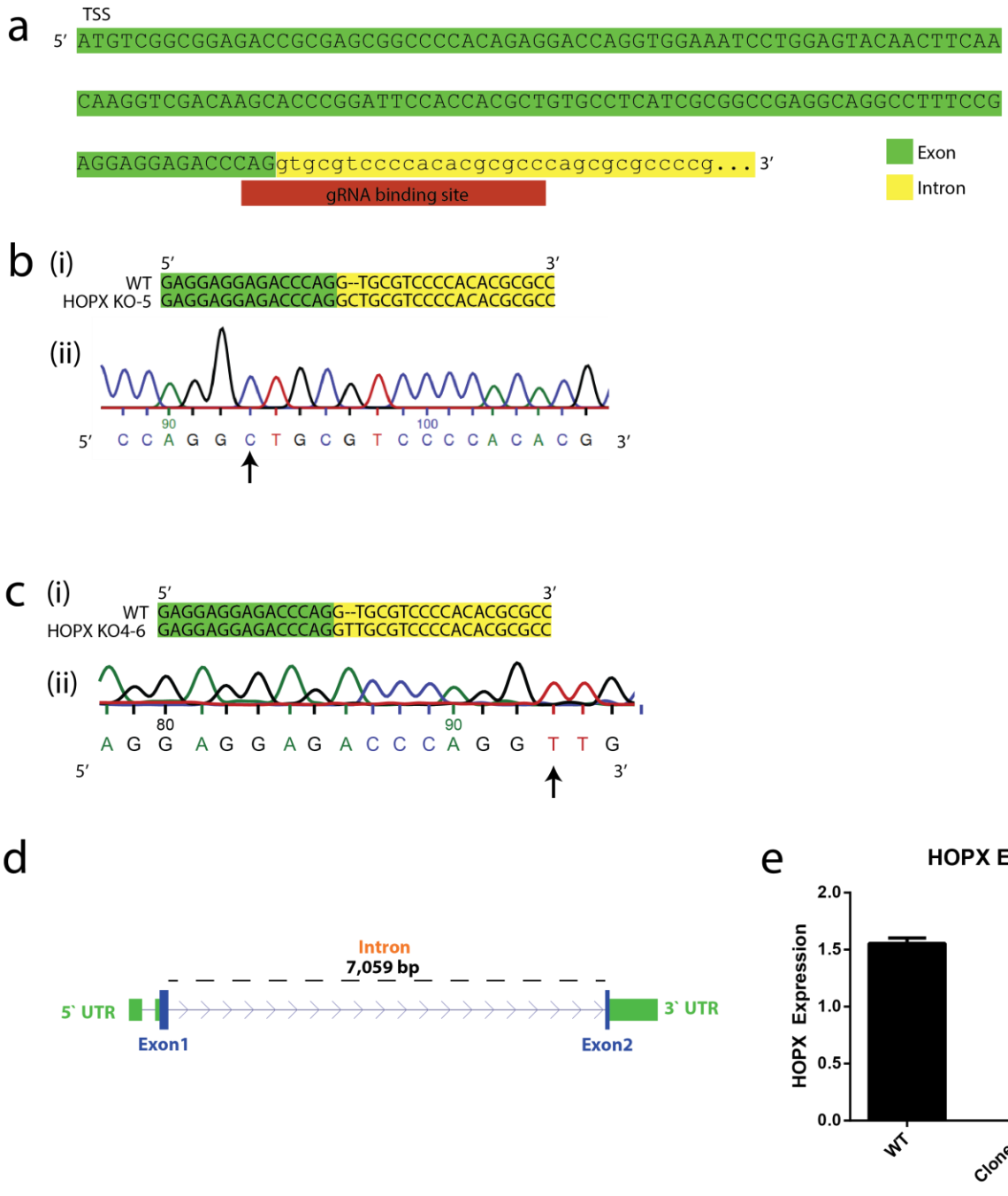

Figure S6 Related to Figure 7. Genotyping of HOPX knockout hESCs. (a) WT HOPX locus nucleotide map from the TSS into the first intron and the gRNA binding site. (b-c) Two different HOPX mutant lines showing (i) sequencing vs the WT and (ii) the raw chromatogram reads. (d) Schematic of HOPX locus used to design intron spanning primers for qRT-PCR. (e) Expression of HOPX transcript in WT vs HOPX clone 5 and 4-6.

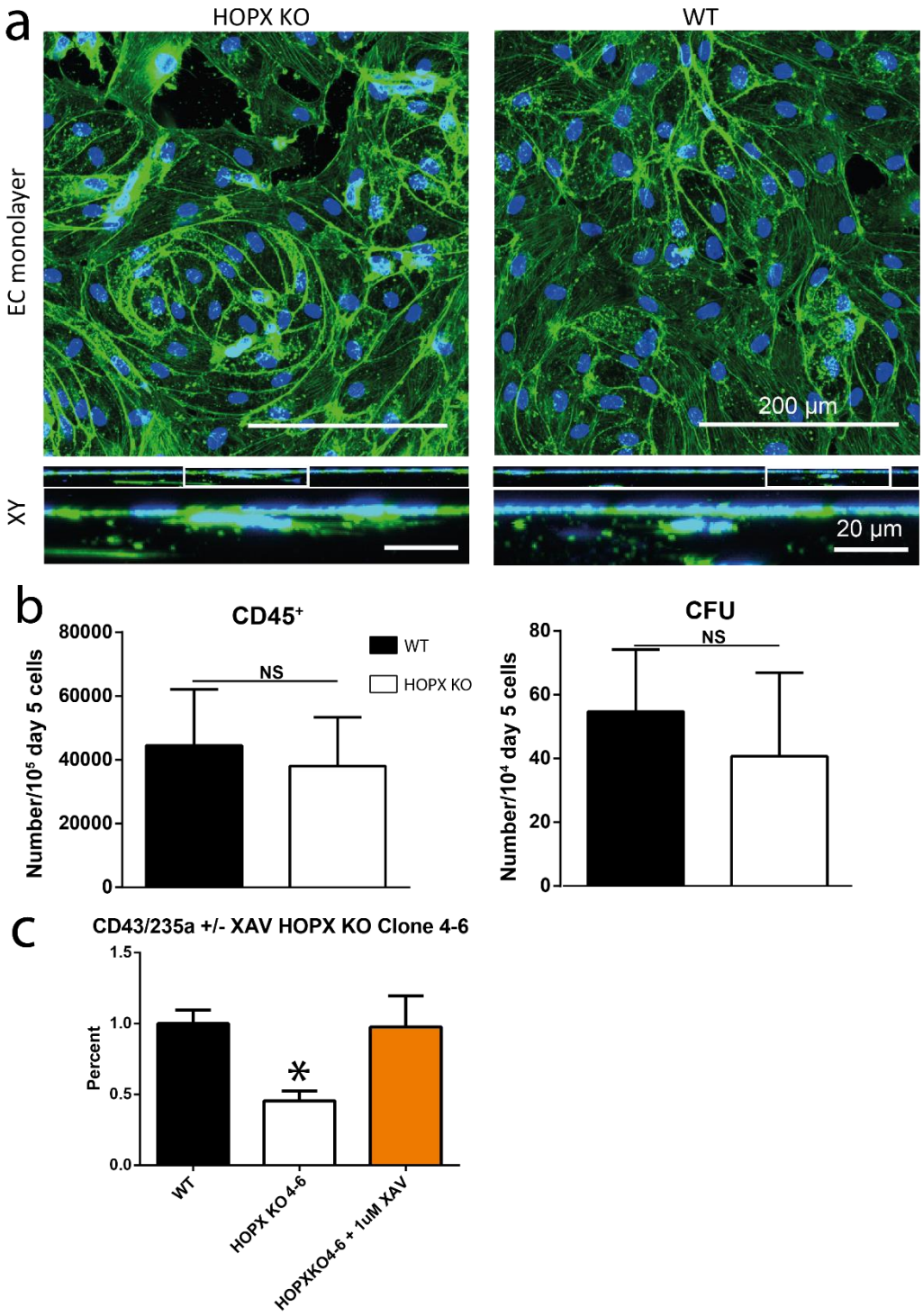

Figure S7 Related to Figure 7. HOPX Loss of function analysis. (a) Images of HOPX KO and day 14 WT H-ECs plated as a monolayer on collagen (top) and assessed for angiogenic activity analyzed by confocal z-stack analysis (below). (b) Definitive hematopoiesis assays after co-culture with OP9 stroma showing CD45+ cell production (left) and CFU (right) in WT and HOPX KO cells. (c) Differentiation of HOPX 4-6 mutant and WT cells into H-ECs showing that addition of the Wnt inhibitor XAV-939 fully rescues hematopoietic deficiencies observed in HOPX mutant cells. n≥3 per group. Values are presented as mean ± sem. *P<0.05 vs. WT.

Table S1 related to Figure 1e. List of lineage regulators and structural molecules used for generating epigenetic profiles in Figure 1e.

Cardiac Regulators	Cardiac Structural	Endothelial Regulators	Endothelial Structural
NKX2-5	MYH6	SOX18	GJA4
GATA4	TNNT2	NR2F2	PECAM1
HAND2	TTN	GATA2	IL16
TBX5	MYL4	FOXC2	ICAM2
ISL1	MYL7	TAL1	CD93
MESP1	MYBPC3	RUNX1	GPR116
MEF2C	DES	NOTCH1	
CELF2	CKM	EGF	
BMP7	GJA5	LMO2	
SMYD1	TNNI1	GATA1	
BMP2	PLN	CD34	
IRX2		LYL1	
		HLX	

Table S10 Related experimental procedures. Quantitative PCR Primers to amplify human genes

Group	Gene	Forward Primer	Reverse Primer
CPC	CELF2	GTGAAAAGTCCAACGCTGTG	CCAGGTGGCAGTGTTGAGC
	GUCY1A3	GAAGGCTGATGAACAGGAGA	TGAGGTCCATAACCCTTGAA
	SHISA2	GCATCGGCTTCCAGTGTCC	CCAACAATGAGGAACGGCAC
	NTF3	GATAAACACTGGAACTCTCAGTGCAA	GCCAGCCCACGAGTTTATTGT
	NXPH2	GAGCCATTGGCAAGAACTAA	CTTCAAATTCCACCACCTTG
C-EC/H-EC	LYL1	TCACCCCTTCCTCAACAGTGT	CGGGCCACCTTCTGGG
	HLX	CAGTTCAGCATCAGTTCCAAGACAC	TCCGGCTTGGTCACGTACTTC
	HOPX	CAAGGTCGACAAGCACCCGGATTC	CATCTCCTTAGTCTGTGACGGA
	PPP1R16B	TTGCTTGCTGTCAACTCGGA	TCTGCCATCTGCATCAGTGT
	CPNE8	TTTGGTGCAAAACTGCCTCC	CCCCTCAATGCCATCACAGT
C-EC	NR2F2	CAGCTGCCTCAAGGCCATAGT	GAAGGGAGGCGAAGCAAAAG
	VAV3	CAGTTTCCATACAAGGAGCCA	CACAGAAGTCATACCGAGCGA
	SOX10	CCCACACTACACCGACCAG	GGCCATAATAGGGTCCTGAGG
	BTG2	AGCCTCATGGTCTCATGCTT	AGACAGGCCTGCTCAACAGT
	TRIM24	CTGAGGTTCCCAGCAGTACA	ACACGTCTTGCAGAGCCATT
H-EC	LIMCH1	CTTCTCCGAGGCGCAGAA	CTGTCCGAAAATCTTTATCACCAA
	NR5A2	CTCATCCGAGCCAATGGACTT	CAAGGCAGCATGGTTCAGA
	GNA14	CAAGAAGCCAAAGCTGACAT	TCAGAAAGGCACAGCATACA
	NEFH	GGAGGCACTGAAAAGCACCAA	CAGGAGTTTTCTGTAAGCGGCT
	DLL4	CCCTGGCAATGTACTTGTGAT	TGGTGGGTGCAGTAGTTGAG
CPC/C-EC	LHX1	TGTAAATGCAACCTGACCGA	ACATCATGCAGGTGAAGCAG
	LRP2	AGACTGGTTCTAACGCCTGTAATC	GCTCTGTGGGTGGTTCATTGG
	EPHA4	ACACGAAAGGGGACCTGGCA	CGCATTGCCTGGACACTGCTCA
	ROR1	CAACAAGAAGCCTCCCTAATGG	CCTGAGTGACGGCACCTAGAA
	DYRK2	GGGGAGAAAACGTCAGTGAA	TCTGCGCCAAATTAGTCCTC

Gene Name	Forward Primer	Reverse Primer
HPRT	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT

NKX2.5	CCAAGGACCCTAGAGCCGAA	ATAGGCGGGGTAGGCGTTAT
SCL	AAGGGCACAGCATCTGTAGTCA	AAGTCTTCAGCAGAGGGTCACGTA
GATA1	CTTTCAGGTGTACCCATTGC	AAAGTCTCCAGGAAGCTGGT
TROY1	GGAGTTGTCTAAGGAATGTGG	GCTGAACAATTTGCCTTCTG
AXIN2	GCGATCCTGTTAATCCTTATCAC	AATTCCATCTACACTGCTGTC

Table S11 related to Figure 1. Linear regression analysis of RNA-seq and qRT-PCR analysis.

Gene	**R^2**
CELF2	0.99873262
BMP7	0.99889375
CCKBR	0.99972347
GUCY1A3	0.9860652
SHISA2	0.9991983
NR2F2	0.98070377
VAV3	0.89507865
SOX10	0.99973927
BTG2	0.96067168
TRIM24	0.99412548
LHX1	0.86262513
CADM1	0.86298582
LRP2	0.95723735
EPHA4	0.95590298
ROR1	0.9929989
LIMCH1	0.99087076
LRFN5	0.99358484
NR5A2	0.99993718
GNA14	0.99972985
NEFH	0.90922764
LYL1	0.93001812
HLX	0.7128004
CD34	0.95086488
HOPX	0.88477861
PP1R16B	0.96976048
AVERAGE R^2	**0.9514502**

Supplemental Experimental Procedures

Cell Culture. RUES2 human ES cells were maintained as previously described (Paige et al., 2012). In brief, cells were plated on Matrigel (BD) coated plates and maintained in an undifferentiated state with mouse embryonic fibroblast (MEF) conditioned media containing 5 ng/mL hbFGF (Peprotech, 100-18B).

HOPX reporter and KO hESCs. We used CRISPR gene editing technology to genetically modify RUES2 hESCs at the HOPX locus **(Figure S3a-e)**. The targeting construct was generated by amplifying homology arms around the HOPX TSS. 3 different gRNAs were designed using established protocols and gRNA2 was identified as the most efficient for homologous recombination **(Figure S3b)**. The HOPX start codon was fused to tdTomato followed by a floxed PGK-driven puromycin cassette **(Figure S3c).** Gene targeting in RUES2 hESCs was performed as previously described (Gantz et al., 2012) with some modifications. Specifically, RUES2 hESCs were pre-treated with Y-27632 for 1 hour. Cells were harvested in a single cell suspension and nucleofected with 20 μg each of the HOPX targeting vector, gRNA, and CAS9-D10A and re-plated on matrigel coated plates in MEF conditioned media supplemented with 5 ng/mL bFGF and Y-27632. After 48 hours, cells were selected with 1 μg/mL puromycin and maintained under continuous puromycin selection. Clones were picked and expanded for southern blot and PCR analysis **(Figure S3d-e)**. PCR analysis of clones was performed using the following primers Forward: TCAGAAGCTGGTCGAGATTCC and Reverse: CTGGCACTTTTCGGGATAAT which were designed to cross from the inserted DNA across the 3` homology arms into genomic DNA. Further confirmation of gene targeting was performed by Southern blot analysis by digestion of genomic DNA using EcoRI and hybridizing with a 5` genomic probe amplified using the following primers: Forward: TGGGAGAGGGAGGATAAGATGAC and Reverse: CCTTGGGTTAGTTGGCTGAGAATC. Correctly targeted clones were expanded. Cre recombination of the PGK-Puro selection cassette was performed using cell permeant cre recombinase (Excellgen).

HOPX KO cells were generated in the HOPX-tdTomato cell backbone by a second round of exposure to the HOPX gRNA CRISPR/Cas9. Clonally amplified cells were picked and screened by sequencing for mutations. Genotyping was performed using the following primers: Forward: CCACAGAGGACCAGGTGGAAATC, Reverse: TGCCGGGGAGAGTGGAAAAGA. The genotyping results are presented in (**Figure S4)** and show disruption of the splice donor site at the end of the first exon of HOPX that results in intronic read-through into a premature stop codon. We validated complete loss of this splice event by PCR.

Human ESC Cardiac Directed Differentiation. WT RUES2 human embryonic stem cells were used in this study. Undifferentiated cells were maintained in mouse embryonic fibroblast-conditioned media. Standard cardiomyocyte directed differentiation using a monolayer platform was performed with a modified protocol based on previous reports (Laflamme et al., 2007; Lian et al., 2012; Paige et al., 2010). The differentiation set up was initiated by plating undifferentiated hESCs as single cells as described previously (Gantz et al., 2012; Palpant et al., 2013). The cultures were treated with CHIR-99021 (Cayman chemical, 13122) for 24 hours before reaching confluence. Cells were induced to differentiate (designated day 0) by replacing the culturing media with RPMI media (Invitrogen, 11875-119) containing 100 ng/mL activin A (R&D SYSTEMS, 338-AC-050), 1:60 diluted Matrigel (BD), and insulin-free B27 supplement (Invitrogen, 0050129SA). An RPMI media change the following day (17 hours) included BMP4 (R&D SYSTEMS, 314-BP-050), 1 μM CHIR-99021, and insulin-free B27 supplement. On day 3, media was changed to RPMI media with B27 minus insulin containing 1 μM XAV-939 (Tocris, 3748). RPMI containing insulin-free B27 supplement was utilized until

differentiation day 7 in which the media is replaced with RPMI containing a B27 supplement that includes insulin (Invitrogen, 17504044). Subsequent media changes included the insulin-containing supplement.

Human ESC Endothelial Differentiation. Endothelial differentiation involved initiation of differentiation with activin A and BMP4 as described above. Endocardial-like ECs are generated by induction with 100 ng/mL activin A on day 0 and 5 ng/mL BMP4 on day 1. Hemogenic endothelial cells are generated by induction with 50 ng/mL activin A on day 0 and 40 ng/mL BMP4 on day 1. For both EC populations, on day 2 media is changed to Stempro34 (Invitrogen, 10640019) backbone media containing 200 ng/mL VEGF (Peprotech, 100-20), 5 ng/mL bFGF (Peprotech, 100-18B), 10 ng/mL BMP4 (R&D SYSTEMS, 314-BP-050), $4x10^{-4}$ M monothioglycerol, 50 ug/mL Ascorbic Acid, 2mM L-Glutamine (Invitrogen, 25030-081), and pen-strep (Invitrogen, 15140-163)(Palpant et al., 2015). Media was not changed until day 5. To determine HOPX worked through the Wnt/β-catenin pathway, we modified the protocol for H-EC differentiation to add 1μM XAV-939 from days 2-5. In some assays, day 5 C-ECs and H-ECs were passaged for maturation into $CD31^+$ cells. This was performed by re-plating cells in gelatin coated tissue culture flasks with EGM media (Lonza, CC-3124) containing 20 ng/mL VEGF (Peprotech, 100-20), 20 ng/mL bFGF (Peprotech, 100-18B), and 1 μM CHIR-99021 (Cayman Chemical,13122). Cells were maintained until day 9 at which point cells were isolated, analyzed by flow cytometry for CD31 expression.

Endothelial Cell Differentiation and Analysis. WT and HOPX KO H-ECs were plated in gelatin coated tissue culture flasks with EGM media (Lonza, CC-3124) containing 20 ng/mL VEGF (Peprotech, 100-20), 20 ng/mL bFGF (Peprotech, 100-18B), and 1 uM CHIR-99021 (Cayman Chemical,13122). Cells were maintained until day 14 at which point cells were isolated, analyzed by flow cytometry for CD31 expression, and cells then applied to secondary assays. **Tube formation assay:** Type I collagen extracted from rat tails as previously described (Zheng et al., 2012) was dissolved in 0.1% acetic acid at a concentration of 15mg/mL and stored at 4°C. Before use, collagen gel was neutralized and diluted with 1M NaOH, EGM, and 10X M199 (Sigma M0650). Gel was then mixed with day 14 cells at a density of $2x10^6$/mL (final collagen concentration: 2mg/mL), pipetted into 4mm diameter well (10uL volume) of angiogenesis μ slides (Ibidi), and allowed to gel for 20 minutes at 37°C. Cultures were fed with EGM supplemented with 20ng/mL VEGF, 20ng/mL bFGF, and 1uM Chiron (EGM+factors) for 2 days, fixed in 3.7% formaldehyde for 10min, washed in PBS, and stained for CD31 and phalloidin. **Angiogenesis assay:** Day 14 endothelial cells were seeded on top of acellular 2mg/mL collagen disks (4mm D x 2mm H) at a density of 400 cells/mm^2 and fed with EGM+factors for 3 days. Samples were fixed for 10 minutes in 3.7% formaldehyde, washed in PBS, and stained for phalloidin and CD31. Three dimensional z stack images of both angiogenesis and tube formation assays were acquired on a Nikon A1R confocal microscope with 20X objective, 1024x1024 resolution. Maximum intensity projections and orthogonal views were obtained using Fiji software.

Quantitative RT-PCR. For quantitative RT-PCR, total RNA was isolated using the RNeasy Miniprep kit (Qiagen). First-strand cDNA was synthesized was performed using the Superscript III enzyme kit (Invitrogen). Quantitative RT-PCR was performed using Sensimix SYBR PCR kit (Bioline) on a 7900HT Fast-Real-Time PCR System (Applied Biosystems). The copy number for each transcript is expressed relative to that of HPRT. Primers used for quantitative PCR are listed in **Table S10**. Lastly, we evaluated the fidelity of the RNA-seq data (taken from 2 biological replicates) by performing linear regression analysis against qRT-PCR data taken from at least 6 biological replicates for each of the 3 progenitor populations **(Table S11)**. 25 genes were analyzed and showed a mean r^2 value of 0.95±0.01 for all genes indicating a high degree of correlation between RNA-seq and qRT-PCR data. Taken together, these data indicate that transcriptional and epigenetic regulation of cell fates based on RNA-seq data represent these mesoderm progenitor populations with high fidelity.

Colony forming assays and hematopoietic differentiation on OP9 Cells. OP9 feeder cells were seeded in 24-well plates the day prior to co-culture. Day 5 unsorted or $CD34^+$-sorted hESC-derived endothelial cells were seeded onto OP9 cells at $1x10^5$ cells per well in (-MEM (Invitrogen) with 10% FBS (Hyclone), penicillin/streptomycin (Invitrogen), and recombinant cytokines (from R&D Systems): hSCF (50 ng/ml), hTPO (20ng/ml), hIL-6 (20 ng/ml), hIL-3 (20 ng/ml), hFLT3L (20 ng/ml). Following 9-11 days of co-culture, non-adherent/loosely adherent cells were removed by pipetting and passed through a 35 (m cell strainer (BD Falcon) and analyzed by flow cytometry. Day 5 hESC-derived endothelial cells or cells harvested following OP9 co-culture were plated for colony forming-unit (CFU) progenitors in methylcellulose containing human cytokines (H4034, Stem Cell Technologies). Colonies were scored by morphology after 12-14 days as small, primitive erythroid (CFU-EryP), macrophage (CFU-Mac), granulocyte/monocyte/macrophage (CFU-GM), large, burst-forming erythroid (BFU-E), or mixed (CFU-Mix) colonies containing both erythroid and myeloid elements.

Immunofluorescence Cells were fixed with 4% paraformaldehyde, permeabilized in PBS containing 0.025% Triton-X, and blocked in PBS containing 1.5% normal goat serum. Cells were stained with alpha-actinin (Sigma, Clone EA-53; Cat.#A7811, 1:800) and dsRed (Clontech, 632496) followed by secondary staining with AlexaFlour-594 Donkey Anti-Goat (Invitrogen lot #1180089, 1:200) or AlexaFlour-594 Goat Anti-Mouse (Invitrogen lot # 1219862, 1:200). Nuclei were counterstained with DAPI.

Flow Cytometry. BAR-venus and HOPX-tdTomato cells were analyzed for intrinsic Venus fluorescence by FACS. Cell derivatives during differentiation were labeled for flow cytometry using the following antibodies: Human PDGFRα APC (R&D Systems) and Human VEGF R2/KDR PE (R&D Systems), CD31 (BD Biosciences, 555445), CD144 (VE-cadherin) (E-biosciences, 17-1449-42), CD34 (BD Biosciences, 340430), CD73 (BD Pharmingen), and cardiac troponin T (Pierce, MA5-12960) or corresponding isotype controls. For hematopoietic analysis, cells were stained with various combinations of the following monoclonal antibodies for cell surface analysis: APC-conjugated anti-human CD43 (Clone 1G10, BD Pharmingen), PE-conjugated anti-human CD235a (Clone GA-

R2, BD Pharmingen), PECy7-conjugated anti-human CD45 (Clone H130, Biolegend), CD73 (BD Pharmingen), FITC-conjugated anti-human CD34 (BD Pharmingen), and PECy7-conjugated anti-human CD41 (Clone HIP8, Biolegend), or corresponding isotype control antibodies. DAPI was used to exclude dead cells. Cells were analyzed using a BD FACSCANTO II (Beckton Dickinson, San Jose, CA) with FACSDiva software (BD Biosciences). Instrument settings were adjusted to avoid spectral overlap. Data analysis was performed using FlowJo (Tree Star, Ashland, Oregon).

RNA-seq. Total RNA was isolated with RNALater (Qiagen, 76104). For each group, 2 biological replicates were submitted for analysis. Samples were submitted to University of Washington High Throughput Genomic Sequencing Center for isolation and analysis. RNA-seq was performed on poly-A enriched samples using Illumina TruSeq. Differentially expressed genes were classified according to gene ontology using the NIAID Database for Annotation, Visualization and Integrated Discovery (DAVID/EASE, http://david.abcc.ncifcrf.gov/). Reads were aligned with Gsnap version 2013-11-27 (Wu and Nacu, 2010) to version hg19 of the human genome sequence obtained from the UCSC Genome Browser (Karolchik et al., 2014). Strand-specific read counts were generated against gene annotations taken from Ensembl release 72 (Flicek et al., 2013). Reads with more than one reported alignment were not counted. Differential expression was evaluated using DESeq version 1.12.1 (Anders and Huber, 2010). Genes with an FDR adjusted p-value (i.e. "q-value") of less than 0.05 were classified as differentially expressed. Separately, FPKM values were computed using Cufflinks version 2.1.1 (Trapnell et al., 2010) and UCSC gene annotations. HOPX expression in human cell types was acquired from the Gene Expression Omnibus (GEO) from the following accession numbers GDS1402 (HOPX - Various normal pure cell cultures) and GDS3842 (Transcription factor-induced pluripotent stem cells).

Gene-level read counts data from single cell RNA-seq were downloaded from http://gastrulation.stemcells.cam.ac.uk. Read counts were normalized for sequencing depth using size factors calculated with DESeq(Anders and Huber, 2010). Expression heatmap was generated for select genes using the same cluster membership and embryonic stage label from Scialdone *et al* (Scialdone et al., 2016) using only samples from E7.0-E7.75 stages. t-SNE plot of samples from E7.0-E7.75 stages was generated the same way as in Scialdone *et al* with highly variable genes. The expression level of HOPX was overlayed on t-SNE plot with grey-red color gradient. TSCAN (Ji and Ji, 2016) was used to reconstruct a pseudo-temporal trajectory for the expression levels of select developmental genes.

Chromatin immunoprecipitation followed by deep sequencing. Day 5 cardiac progenitor cells, C-ECs, and H-ECs were fixed in 1% formaldehyde for 10 minutes at RT to crosslink the chromatin. Following that glycine was added to a final concentration of 0.125 M. Cells were washed twice with cold PBS and frozen at -80ºC. Crosslinked samples were submitted to the University of Washington High Throughput Genomic Sequencing Center for ChIP-seq analysis for H3K4me3 and H3K27me3 histone modifications. The chromatin was sonicated to small fragments (<500 bp) and histone bound chromatin was immunoprecipitated using antibiodies specific to H3K4me3 (Cat # 9751, Cell Signaling) or H3K27me3 (Cat # 07-449; Millipore). For each IP, Dynabeads (M-280, sheep anti-rabbit IgG, Invitrogen) were incubated with antibodies for 6 hr at 4°C and then incubated overnight with ~100 μg sheared chromatin. The complexes were rinsed sequentially with IP wash buffer I (50 mMTris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA [pH8.0], 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), high salt buffer (50 mMTris-HCl pH8.0, 0.5 M NaCl, 1 mM EDTA [pH8.0], 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), IP wash buffer II (50 mMTris-HCl [pH8.0], 1 mM EDTA pH8.0, 1% NP-40, 0.7% sodium deoxycholate, 0.5 M LiCl), and TE buffer (10 mMTris-HCl [pH8.0], 1 mM EDTA pH8.0). The complexes were incubated with elution buffer (10 mMTris-HCl [pH 8.0], 0.3 M NaCl, 5 mM EDTA [pH8.0], 0.5% SDS) supplemented with RNase A (Ambion) at 65°C overnight. After separation, the DNA was treated with Proteinase K and purified by using PCR purification column (QIAGEN). The Illumina indexed sequencing libraries were prepared following a standard protocol using PE adapters (Illumina). The libraries were used to generate sequencing clusters on flowcells using cBot, and sequenced on Illumina HiSeq 2000 to generate PE 2X36 base pair reads. Each library was on average sequenced to a depth of ~18 million tags that mapped uniquely to the human genome (hg19). Differentially marked genomic regions were identified with diffReps version 1.55.4 (Shen et al., 2013) and annotated to the closest genes. Genes associated with at least one significant genomic region (FDR less than 0.05 and fold change >1) were classified as differentially marked. When a gene is annotated with multiple significant genomic regions, the most significant one is assigned to that gene. Our data has been curated & approved by Gene Expression Omnibus: GEO97080.

ChromHMM and identification of bivalent genes This analysis is similar as previously described by Sachs et al(Sachs et al., 2013). A 9 state segmentation of the genome was generated by ChromHMM v1.10 (Ernst and Kellis, 2012). 4 samples (2 replicates for each ChIP-seq chromatin mark) of each lineage were used. All parameters were ChromHMM defaults except stopping criterion is 500 iterations (-r 500 in LearnModel) and the read count in a ChIP-seq bin must be enriched 3 fold over input samples to be called present (-f 3 in binarizeBed). K-means clustering with k=2 was used to group the state emission parameters of each sample generated by ChromHMM into “on” or “off”. Each of the 9 chromatin states were classified into “bivalent”, “H3K27me3 only”, “H3K4me3 only” or “none” in each cell type based on the following rule: bivalent if it is called present across all 4 samples; “none” if it is called absent in all samples; “H3K27me3 only” if called present in both H3K27me3 replicates but not in both H3K4me3 samples; “H3K4me3 only” if called present in both H3K4me3 replicates but not in both H3K27me3 samples. Using the classification result, each genomic region is transformed from one of 9 ChromHMM states to one of four states (“bivalent”, “H3K27me3 only”, “H3K4me3 only” or “none”). Based on hg19 RefSeq annotation, a gene is called bivalent if there is at least one bivalent genomic region around 5kb of transcription start site (TSS) in any of its transcripts; a gene is called H3K4me3-only if no transcripts were bivalent and if any transcript had H3K4me3-only regions around 5kb of TSS; a gene is called H3K27me3-only if it is neither bivalent nor H3K4me3-only and if any transcript had H3K27me3-only regions around 5kb of TSS. The percentage of genes in each of the 4 categories is similar to previous reports (Sachs et al., 2013).

Sashimi isoform analysis Reads were aligned to version hg19 of the human genome using Gsnap (Wu and Nacu, 2010). Transcript abundance was estimated using a Markov chain Monte Carlo algorithm to sample from a hierarchical Bayesian model. The model includes parameters separately capturing transcription and splicing rates at the sample, condition and experiment level. Under this model, estimates from transcripts with very low data are effectively shrunk towards each other to avoid spurious predictions of large effect-size. This model was run gene models defined in version 75 of the Ensembl database (Flicek et al., 2014). We selected as differentially spliced transcripts with a shift in relative abundance (defined as transcript abundance divided by the sum of the abundances of all transcripts of the same gene) of at least 0.3, with posterior probability > 0.5. Principal component analysis was performed on point estimates of relative abundance using R 3.1.0. Sashimi plots were generated with the sashimi plot routine (arXiv:1306.3466 [q-bio.GN]), which is part of the MISO python package (MISO version 0.5.2 (Katz et al., 2010)). Default plotting parameters were used. Fetal ventricle and hESC-derived cardiomyocytes aged to one year were generated as previously published(Kuppusamy et al., 2015).

Generating regulator lists The basic principle of generating a putative regulator list is to identify genes with significant changes in levels of expression and H3K27me3 or H3K4me3 marks. This includes both filtering and ranking steps. First, only genes with significant changes in levels of expression and at least one histone marks (H3K27me3 or H3K4me3) in a pairwise lineage comparison are kept for further analysis. For each gene *g* in each data set *D* in each pairwise comparison of two cell types *i* and *j*, score S is defined as:

$$S_{D(i:j;g)} = sign\left(log2\left(\frac{d_{i;g}}{d_{j;g}}\right)\right) * (-log10(FDR_{i:j;g}))$$

where *i* and *j* = C-ECs, H-ECs or CPCs; $d_{i;g}$ and $d_{j;g}$ the values of gene *g* in data set *D (expression, H3K27me3, or H3K4me3)*, and $FDR_{i:j;g}$ the false discovery rate of gene *g* in the pairwise comparison between lineages *i* and *j*. S is a signed score based on the statistical significance of change for genes with higher values in lineage *i*.
Then the three data sets are integrated such that genes with consistent changes (higher expression, higher active H3K4me3 marks and lower repressive H3K27me3 marks or vice versa) in either direction will get the highest absolute scores:

$$S_{i:j;g} = S_{Expression(i:j;g)} + S_{H3K4me3(i:j;g)} - S_{H3K27me3\ (i:j;g)}$$

To identify regulators for a single lineage *i*, only genes with positive scores from pairwise comparisons involving lineage *i* are kept, and the combined scores are defined as:

$$S_{i;g} = S_{i:j;g} + S_{i:k;g}$$

To identify shared regulators for two lineages *i* and *j*, only genes with positive scores from pairwise comparison involving a third lineage *k* are kept, and the combined scores are defined as:

$$S_{(i:j);g} = S_{i:k;g} + S_{j:k;g}$$

Genes with FDR values smaller than 1e-15 were truncated at 1e-15 to prevent expression data from dominating the integrative significance score calculation.

The R code to generate integrated, expression only and chromatin only lists are deposited on Github:
https://github.com/yuliangwang/regulators_integrating_RNAseq_ChIPseq

Calculating H3K4me3 breadth H3K4em3 peaks were identified using MACS version 1.4.2. Two replicates of the same cell type were combined. Default parameters were used. 18109, 39445 and 37880 peaks were identified for CPCs, C-ECs and H-ECs, respectively. Peak length and peak summit position and peak height were produced by MACS by default. Peaks were assigned to the nearest genes by the ChIPpeakAnno package(Zhu et al., 2010) from Bioconductor(Gentleman et al., 2004). If multiple peaks were assigned to the same gene, the peak with most significant false discovery rate and highest fold enrichment over background was selected. Distance from peak to transcription start site (TSS) was produced by ChIPpeakAnno package by default.

Statistics. Single variable analysis between 2 samples was compared by Student's t-test. Results are presented as mean ± SEM. For all results: * $P < 0.05$.

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.
Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.
Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., *et al.* (2013). Ensembl 2013. Nucleic Acids Res *41*, D48-55.
Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., *et al.* (2014). Ensembl 2014. Nucleic Acids Res *42*, D749-755.

Gantz, J.A., Palpant, N.J., Welikson, R.E., Hauschka, S.D., Murry, C.E., and Laflamme, M.A. (2012). Targeted genomic integration of a selectable floxed dual fluorescence reporter in human embryonic stem cells. PLoS One *7*, e46971.
Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5*, R80.
Ji, Z., and Ji, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. Nucleic Acids Res *44*, e117.
Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., *et al.* (2014). The UCSC Genome Browser database: 2014 update. Nucleic Acids Res *42*, D764-770.
Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods *7*, 1009-1015.
Kuppusamy, K.T., Jones, D.C., Sperber, H., Madan, A., Fischer, K.A., Rodriguez, M.L., Pabon, L., Zhu, W.Z., Tulloch, N.L., Yang, X., *et al.* (2015). Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes. Proc Natl Acad Sci U S A *112*, E2785-2794.
Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S., *et al.* (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nat Biotechnol *25*, 1015-1024.
Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proc Natl Acad Sci U S A *109*, E1848-1857.
Paige, S.L., Osugi, T., Afanasiev, O.K., Pabon, L., Reinecke, H., and Murry, C.E. (2010). Endogenous Wnt/beta-catenin signaling is required for cardiac differentiation in human embryonic stem cells. PLoS One *5*, e11134.
Paige, S.L., Thomas, S., Stoick-Cooper, C.L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G., *et al.* (2012). A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell *151*, 221-232.
Palpant, N.J., Pabon, L., Rabinowitz, J.S., Hadland, B.K., Stoick-Cooper, C.L., Paige, S.L., Bernstein, I.D., Moon, R.T., and Murry, C.E. (2013). Transmembrane protein 88: a Wnt regulatory protein that specifies cardiomyocyte development. Development *140*, 3799-3808.
Palpant, N.J., Pabon, L., Roberts, M., Hadland, B., Jones, D., Jones, C., Moon, R.T., Ruzzo, W.L., Bernstein, I., Zheng, Y., *et al.* (2015). Inhibition of β-catenin signaling respecifies anterior-like endothelium into beating human cardiomyocytes. Development.
Sachs, M., Onodera, C., Blaschke, K., Ebata, K.T., Song, J.S., and Ramalho-Santos, M. (2013). Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep *3*, 1777-1784.
Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N.K., Macaulay, I.C., Marioni, J.C., and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. Nature *535*, 289-293.
Shen, L., Shao, N.Y., Liu, X., Maze, I., Feng, J., and Nestler, E.J. (2013). diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PLoS One *8*, e65598.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.
Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873-881.
Zheng, Y., Chen, J., Craven, M., Choi, N.W., Totorica, S., Diaz-Santana, A., Kermani, P., Hempstead, B., Fischbach-Teschl, C., López, J.A., *et al.* (2012). In vitro microvessels for the study of angiogenesis and thrombosis. Proc Natl Acad Sci U S A *109*, 9342-9347.
Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.