

SUPPLEMENTAL INFORMATION CONTENTS

Expanded Material & Methods

Supplemental Figs. S1-S16

- Fig. S1. Minimal batch effects of scRNA-seq samples.
- Fig. S2. Identification of major cell types within scRNA-seq samples.
- Fig. S3. Embryonic EC clusters and exclusion of a red blood cell cluster.
- Fig. S4. Selection of specific marker genes for EC clusters.
- Fig. S5. Determination of tissue-specific heterogeneity of embryonic ECs by marker genes.
- Fig. S6. Identification of embryonic tissue-specific EC marker genes with FACS-qPCR and scRNA-seq.
- Fig. S7. Identification of specialized endothelial clusters.
- Fig. S8. Quality control of MERFISH study.
- Fig. S9. In silico selection of endothelial cells (ECs) and smooth muscle cells (SMCs) from MERFISH data.
- Fig. S10. Correspondence of EC clusters between MERFISH and scRNA-seq datasets.
- Fig. S11. Spatial distribution of subpopulations of highly heterogeneous EC clusters.
- Fig. S12. Spatial verification of embryonic tissue-specific EC marker genes.
- Fig. S13. Spatial expression pattern of selected marker genes.
- Fig. S14. Spatial correlation of ECs and SMCs.
- Fig. S15. Characterization of arterial and venous differentiation trajectories.
- Fig. S16. Knockout of selected transcription factors.

Supplemental Tables S1-S8

- Table S1. Major resources table.
- Table S2. Key quality metrics of the sample libraries and sequencing process.
- Table S3. Marker genes used in heatmaps. (Provided as excel file)
- Table S4. Top markers for EC clusters. (Provided as excel file)
- Table S5. Top enriched GO terms. (Provided as excel file)
- Table S6. Expression of transcription factors shown in Figure 6C. (Provided as excel file)
- Table S7. Oligonucleotides used in this study.
- Table S8. Gene code book for MERFISH study. (Provided as excel file)

EXPANDED MATERIALS AND METHODS

Resources used in this study are listed in Table S1. Oligonucleotide sequences are provided in Table S7.

Mice

All relevant ethical regulations for animal testing and research were followed. All procedures were performed following protocols approved by the Boston Children's Hospital Institutional Animal Care and Use Committee. C57BL6 mice were maintained at 65-75°F with 40-60% humidity on a normal 12-hour light, 12-hour dark cycle, and were fed a standard diet. Cdh5(PAC)-CreERT2, R26^{TdTomato}, R26^{fSTRAP} and R26^{mTmG} mice were described previously⁴⁶⁻⁴⁹. For scRNA-seq experiments, E9.5, E12.5 and E15.5 embryos were used. For MERFISH experiments, E9.5 and E15.5 embryos were used. For FACS-qPCR experiments, E12.5 and E15.5 embryos and 10-week-old adult mice were used. We used multiple embryos from the same mother for one sample so the sample should contain both sexes and our adult samples also contained both sexes.

Embryo dissociation and EC enrichment for scRNA-seq

The genotype information about all scRNA-seq samples is listed in the Fig. S1A. Timed pregnancy was determined by vaginal plug checking and this was performed for all pregnancies where a gestational age is indicated. Tamoxifen (0.05 mg/g body weight by gavage) was administered at E7.5, E8.5, E10.5, or E12.5. Embryos were harvested at E9.5, E12.5, E15.5, and embryos from the same stages were pooled together. The tail, limbs, and liver were removed from E12.5 and E15.5 embryos. The remaining tissue or whole E9.5 embryos were cut into small pieces using sterile scalpels and washed three times with Hank's Balanced Salt Solution (HBSS) to remove most blood cells. Tissues were incubated in 2 mg/ml Collagenase Type II (Worthington, Cat# LS004174) and 50% (v/v) Accutase (StemCell Technologies, Cat# 07920) solution for 15 mins at 37°C. The sample was further treated with gentleMACS Dissociator (Miltenyi Biotec, Germany) twice to get single-cell suspensions. The enzyme was neutralized using 2% FBS -HBSS. The cell suspension was filtered through a 40-µm cell strainer (Corning). After centrifugation (300g, 5 min, 4 °C) the supernatant was removed and the pellet was resuspended in flow buffer (1x phosphate-buffered saline buffer [PBS], 0.5% bovine serum albumin) and then incubated with Rat anti-human/mouse CD31-FITC (Thermo Fisher, Cat# 11-0311-82) for 15 mins at room temperature. Magnetic-Activated Cell Sorting (MACS) was performed with EasySep™ FITC Positive Selection Kit II (StemCell Technologies, Cat# 17682) to enrich ECs (Fig. S1B). Selected ECs bound to microbeads were retained in microfuge tubes using a magnetic rack so that they could be easily eluted and had a high viability. To achieve high viability, all incubation steps were performed only once, resulting in relative lower EC purity. The enriched ECs were resuspended in DMEM + 10% FBS. Viability and cell number were measured using aCountess Automated Cell Counter (Thermo Fisher).

scRNA-seq library preparation and sequencing

Single-cell suspension with viability > 85% were adjusted to 1E6 cells per milliliter and loaded on the Chromium Controller (10x Genomics) with a targeted cell recovery of 7,000-10,000 per

reaction. Cellular samples with viability < 85% were discarded before cell capture step. 3' gene expression libraries were prepared according to the manufacturer's instructions of the v2 or v3 Chromium Single Cell Reagent Kits (10x Genomics). Quality control of cDNA and final libraries was done using Bioanalyzer High Sensitivity DNA Analysis (Agilent) or 4200 TapeStation System (Agilent). Poor quality (wrong size, low yield) cDNAs and libraries were excluded. Libraries that passed quality control were sequenced using NextSeq 500 (Illumina) at Harvard Medical School.

scRNA-seq data processing and in silico EC selection

Single cell RNA-seq data from the 10x Genomics platform were processed with CellRanger software (version 7.0.0, 10x Genomics). Data were mapped to the mm10 reference genome. After doublet filtering by Scrublet (RRID:SCR_018098), the data were analyzed using the Seurat package⁵⁰. Cells meeting these criteria were removed: (i) < 500 genes (low quality) or > 5,000 genes detected (potential doublets); (ii) >10% of unique molecular identifiers (UMIs) originating from the mitochondrial genome. Data normalization and batch effect removal was performed using Seurat. For in silico EC selection, we first identified highly variable genes using the Seurat's FindVariableGenes function (mean lower threshold = 0.0125, mean higher threshold = 3, dispersion threshold = 0.5). Highly variable genes were then integrated by Seurat's IntegrateData function, summarized by principal component analysis (PCA) using the RunPCA function, and visualized using Uniform Manifold Approximation and Projection (UMAP; top 20 principal components). Graph-based clustering was performed to cluster cells according to their gene expression profile using the FindClusters function in Seurat (clustering resolution = 0.6, k-nearest neighbors = 20). EC clusters were annotated based on the expression of known EC and non-EC marker genes, including *Etv2* (early EC), *Pecam1* and *Cdh5* (EC), *CD45*, *Lyz*, *Ly86* (white blood cells), *Hba-a1*, *Hba-x*, *Hba-y* (red blood cells), *Epcam*, *Cdh1*, *Krt18* (Epithelial cells), *Map2*, *Tubb3*, *Hes5*, *Pax3* (neural cells). Contaminating cell clusters (non-ECs) were removed, and all subsequent analysis was performed on ECs only.

Data visualization

Seurat was used for UMAP, violin plot and bar graph visualization. Expression values are scaled, normalized, and natural-log transformed values generated by Seurat. The R package 'pheatmap' (RRID:SCR_016418) was used for heatmap visualization. All heatmaps were based on cluster-averaged gene expression to account for cell-to-cell transcriptomic stochastics, and data was autoscaled for visualization. The R package ggplot2⁵¹ was used to generate bar and scatter plots. Data matrices for each heatmap can be found in Data S1 in the Supplement.

SCENIC

We used all cells from each sample as input to carry out transcription factor network inference. Analysis was performed using the R package SCENIC³⁷ (version 1.1.0, which incorporates RcisTarget 1.2.0 , AUCell 1.4.1, and RcisTarget mm9 motif databases). Activity of the regulatory networks was evaluated on the full dataset in the scoring step with AUCell. Regulons annotated as "extended" include target genes harboring motifs that have been linked to the respective transcription factor by lower confidence annotations.

GO enrichment analysis

We used the Seurat FindAllMarkers function and FindMarkers function to identify cluster feature genes among multiple clusters and differentially expressed genes (DEGs) between two given clusters (only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25), respectively. Network enrichment analysis was performed using Metascape (<http://metascape.org>). Gene ontology analysis was performed using clusterProfiler (RRID:SCR_016884). The analysis used the org.Mm.eg.db database (ont = "BP"). GO terms with Qvalue <0.05 in one cluster were kept. Top 20 GO terms are listed in Table S5 in the Supplement.

Transcription factors

Transcription factors used in Figure 6C were the 1665 human transcription factors downloaded from Animal TFDB (<http://bioinfo.life.hust.edu.cn/AnimalTFDB/>).

Arterial and venous feature score

The arterial and venous feature scores were calculated as described⁷, based on 13 previously reported aEC marker genes (*Bmx*, *Cxcr4*, *Dll4*, *Efnb2*, *Epas1*, *Gja4*, *Gja5*, *Hey1*, *Igfbp3*, *Mecom*, *Nrp1*, *Unc5b* and *Vegfc*) and 3 previously reporter vEC marker genes (*Aplnr*, *Nr2f2* and *Nrp2*). We scaled the $\log_2(\text{TPM}/10 + 1)$ expression values of each of the 20 marker genes to 0-10 scale among all the sample cells after quality control. Second, for each cell, we averaged the scaled values of arterial genes and venous genes, respectively. Third, the averaged values were rescaled to 0–10 scale across all the sample cells to yield the arterial and venous score, respectively. For each population, the arterial and venous scores of the constituent cells were averaged, and a 50% confidence ellipse was calculated to show the main distribution range.

Single cell trajectories

To infer the trajectory of aEC and vEC differentiation, we used 3 methods, Slingshot³⁴, URD³⁵ and scVelo⁵². We excluded subclusters that were less relevant to arteriovenous differentiation or were tissue-restricted. Variable genes identified using the Seurat FindVariableGenes function were used for trajectory analysis. Cluster labels produced by our clustering procedures were used as landmark cluster assignments of individual cells. For Slingshot, the analysis was performed using the following parameters: reducedDim = "UMAP", start.clus = "early_EC", end.clus = c("artery-2", "vein") and stretch = 0. For visualization, clusters were ordered based on their pseudotime value. For URD, the root cells were fixed to the "early_EC", which was the most anterior cluster based on manual annotation. The pseudotime density distribution was calculated using the floodPseudotime and floodPseudotimeProcess functions. Potential tip populations were determined and refined. The buildTree function was then performed to determine potential relationships between cell populations (divergence.method = "preference", cells.per.pseudotime.bin = 25, bins.per.pseudotime.window = 8, save.all.breakpoint.info = T, p.thresh=0.001, tip.use = c("artery-2", "vein", "plexus-2", "capil-lung")). When performing the RNA velocity analysis, we first used velocito³⁶ to recount the spliced reads and unspliced reads based on previously aligned bam files, then the dataset was assessed by a steady-state deterministic model using scVelo.

Dynamically changed genes along arterial and venous ECs trajectories

Genes changed along the arterial and venous paths were identified by Slingshot using ‘wilcox’ method $P<0.05$ and $\text{logfc} > 0.25$ thresholds. If these criteria yielded more than 50 genes, only the top 50 genes ranked by logfc were used. For visualization of changes in gene expression across EC populations, we inferred the pseudo-order of cells involved in each lineage. All cells within the given clusters were included for analysis. In the visualizations, cells are ordered by pseudo-order, and scaled by row. Genes were assigned to expression patterns based on the cluster with maximal expression during the differentiation trajectory and then sorted by descending expression level within that cluster.

Tissue-specific EC sorting

Cdh5(PAC)-CreERT2 mice were bred with *R26^{TdTomato}* mice. Tamoxifen gavage was performed at E8.5, E10.5 and E12.5, and 0.12 mg/g body weight of tamoxifen was introduced. Brain (n = 4), heart (n = 4), lung (n = 3), liver (n = 3), kidney (n = 3), and small intestine (n = 2) were obtained from adult mice. The same tissues were obtained from E15.5 embryos (brain, n = 3; heart, n = 2; lung, n = 4; liver, n = 3; kidney, n = 2; small intestine, n = 2). At E12.5 stage, only the brain (n = 3), heart (n = 3) and liver (n = 3) were harvested. At embryonic stages, 3 ~ 4 embryos were pooled together to make one sample. Dissected tissues were dissociated into single-cell suspensions using Collagenase Type II (Worthington, Cat# LS004174) and 50% (v/v) Accutase (StemCell Technologies, Cat# 07920) and washed with PBS supplemented with 0.5% bovine serum albumin (BSA). The TdTomato^+ ECs and TdTomato^- non-ECs were sorted using a Sony SH800 Cell Sorter (Sony, Tokyo, Japan).

Basement membrane matrices

To prepare culture plastics for hiPSCs, Geltrex (Thermo Fisher) was diluted 1:100-1:200 in cold DMEM (Thermo Fisher) and was used to coat tissue culture plastics at 37°C for 1 hour. Recombinant human truncated vitronectin (Thermo Fisher, Cat # A14700) was diluted to a 10 mg/mL stock in cold PBS (lacking Ca^{2+} or Mg^{2+}) and was used to coat tissue culture plastics at 37°C for at least 1 hour for hpSC-EC differentiation. 0.1% gelatin solution (Millipore) was used to coat tissue culture plastics at 37°C for at least 1 hour for ECs culture.

Human pluripotent stem cell lines

All cells in this study were cultured in standard incubator conditions: 5% CO_2 , 21% O_2 , and 37°C. The following human induced pluripotent stem cell lines were used in this study: wild-type PGP1, and *MECOM* KO, *USF1* KO, *XBP1* KO, and *IRX5* KO PGP1 hiPSCs. Undifferentiated hiPSCs were propagated in mTeSR1 medium (StemCell Technologies) + 1% penicillin/streptomycin (Thermo Fisher) on Geltrex basement membrane matrix-coated plates (described below). mTeSR1 medium was changed daily. To maintain cultures of undifferentiated hiPSCs, they were passaged when they became ~80% confluent by treating them for 7 minutes with Versene (Thermo Fisher) at room temperature. Subsequently, Versene was removed, mTeSR1 was added, and then hiPSCs were seeded onto new plates that had been precoated with Geltrex basement membrane matrix (described below) in mTeSR1 medium + 1% penicillin/streptomycin + 1 μM thiazovivin (StemCell Technologies).

hiPSC differentiation into arterial and venous ECs

hiPSCs were differentiated into arterial and venous ECs as described previously.⁴⁰ All standard hiPSCs cultures were grown in mTeSR1 medium on Geltrex coated plates. One day prior to differentiation (Day -1), 80% - 90% confluent hiPSCs were dissociated into single cells with Accutase (StemCell Technologies) and reseeded on Geltrex-coated plates at a density of 0.2-0.5x10⁵ cells/cm² in mTeSR1 medium supplemented with 1 µM thiazovivin (Rock Inhibitor). Day 0 hiPSCs were differentiated to mid primitive streak in Chemically Defined Medium 2⁴⁰ (CDM2) supplemented with Activin A (30 ng/mL, R&D Systems), BMP4 (40 ng/mL, R&D Systems), CHIR99021 (6 mM, Tocris), FGF2 (20 ng/mL, Thermo Fisher). On Day 1, the differentiation media were replaced by fresh lateral mesoderm induction media [CDM2 media + BMP4 (40 ng/mL), GDC-0941 (2.5 mM, Collagen Technology), Forskolin (10mM, Tocris), SB-505124 (2 mM, Tocris), VEGF (100 ng/mL, R&D Systems), XAV939 (1 mM, Tocris) and ascorbic acid-2-phosphate (AA2P; 200 mg/mL, Sigma)]. On Day 2 - 3, the differentiation media were replaced by fresh aDiff differentiation medium [CDM2 media + Activin A (15 ng/mL), DMH1 (250 nM, Tocris), GDC-0941 (2.5 mM), VEGF (100 ng/mL), XAV939 (1 mM), AA2P (200 mg/mL)], or vDiff differentiation medium [Day 2: CDM2 media + SB505124 (2 mM), DMH1 (250 nM), RO4929097 (2 mM, Collagen Technology), VEGF (100 ng/mL), XAV939 (1 mM), AA2P (200 mg/mL); Day 3: CDM2 media + SB505124 (2 mM), RO4929097 (2 mM), PD0325901 (500 nM, Tocris), CHIR99021 (1 mM) and AA2P (200 mg/mL)]. Day 4, flow cytometry experiments were performed to determine differentiation efficiency. Differentiations in which WT groups yielded < 10% a/vECs were considered unsuccessful. These experiments were excluded from analysis.

Establishment of gene knockout hiPSC lines

To establish knockout (KO) hiPSC lines for *USF1*, *XBP1*, *MECOM*, and *IRX5*, we adopted a previously described strategy⁵³. Briefly, a double-stranded donor fragment that constitutively expresses the blasticidin resistance gene was introduced into the targeted site by two guide RNAs, recombination-directing single-stranded oligodeoxynucleotides (ssODNs), and Cas9 protein, so that PGP1 cells in which the gene-of-interest was knockout by the inserted cassette could be selected by blasticidin. We used the Alt-R CRISPR-Cas9 system [Integrated DNA Technologies (IDT)] to do the genome editing. In brief, gRNA was prepared by combining crRNA and tracrRNA (IDT, Cat# 1072533) to a final duplex concentration of 100 µM. Ribonucleoprotein (RNP) complex was prepared by mixing 4 µl of Cas9 protein stock (IDT, Cat# 1074181) and 5 µl of freshly prepared gRNA. After incubation at room temperature for 15 min, the mixture was then combined with Alt-R Cas9 Electroporation Enhancer (IDT, Cat# 1075915), ssODNs and donor fragments containing the blasticidin resistance gene. The final mixture was delivered into hiPSCs using the Human Stem Cell Nucleofector Kit (Lonza #VPH-5012, program B-016). After 24 hours, blasticidin selection (30 ~ 50 µg/ml) was performed. Six days after electroporation, cells were dissociated and re-plated onto a 6-well plate in a single cell format. Single cell-derived colonies were transferred to 24-well plates. Clone genotype was determined by genotyping PCR and Sanger sequencing of genomic DNA. Sequences of primers, gRNAs, ssODNs and HDR donors used in this study are provided in Table S7 in the Supplement.

Flow cytometry

For flow analysis experiments, cells were dissociated into single-cell suspensions using Accutase (StemCell Technologies, Cat# 07920) and washed with PBS supplemented with 0.5% BSA and 0.2 mM EDTA. Cells were stained with flow cytometry antibodies and analyzed using a BD LSRFortessa Cell Analyzer (BD Biosciences, NJ, US) and FlowJo software (Tree Star Inc., Ashland, OR). Antibody incubation was performed at room temperature for 20 minutes followed by three washes with PBS buffer. Isotype antibodies were used for control samples for PE/FITC/APC/BV421-conjugated antibodies, and no antibody controls were used for PE-Cy7/PerCP-Cy5.5-conjugated antibodies. For data analysis, cells were sorted using forward and side scatter area parameters, followed by height and width parameters to discriminate between single and double cells. Next, singlets were preceded by marker analyses and population frequency calculations. Experiments where wild-type groups showed unsuccessful differentiation (below 10% efficiency) were omitted from the final statistical analysis.

Microscopy

Fluorescent images were taken using a FV3000 Confocal Laser Scanning Microscope (Olympus) and a Keyence BZ-X700 Fluorescence Microscope (Keyence).

Quantitative reverse transcription PCR

Cultured cells and purified cells were treated with TRIzol Reagent (Thermo Fisher, Cat# 15596026) and total RNA was then isolated with an RNA Clean & Concentrator Kit (ZYMO Research, Cat# ZR1013). The genomic DNA was removed by the DNase I set from the same kit. cDNA was prepared with the SuperScript III First-Strand Synthesis System (Thermo Fisher, Cat# 18080051), according to the manufacturer's instructions. Quantitative PCR was performed using Power SYBR Green PCR Master Mix (Thermo Fisher, Cat# 4368708), and detection was achieved using the CFX384 Real Time PCR System thermocycler (Biorad). General PCR conditions were used for all detected genes: 95°C, 15s; 60°C, 60s; 40 cycles. Expression of target genes was normalized to *Actin Beta* (*ACTB* or *Actb*). Quantitative PCR primer sequences are listed in Table S7 in the Supplement.

RNAscope mRNA in situ hybridization assay

Fresh embryos were fixed in freshly prepared 4% paraformaldehyde (Thermo Fisher) for 24 hours at 4°C. Fixed embryos were soaked in graded concentrations of sucrose/PBS (10%, 20%, 30%). Sections were cut at 10-μm thickness using a microtome and placed onto SuperFrost Plus slides (Thermo Fisher). Fixed-frozen tissue slides were processed according to the protocol of RNAscope Multiplex Fluorescent Assay v1 or v2 (ACDBio). In brief, slides stored at -80°C were first washed with PBS and baked for 30 min at 60°C in the RNAscope Hybridization oven (ACDBio), followed by a 15 min fixation in 4% paraformaldehyde. Then dehydration with ethanol and hydrogen peroxide application were performed, followed by target retrieval, which was performed with RNAscope Target Retrieval Reagents (ACDBio, Cat# 322000) in an Oster steamer as recommended by ACD. After target retrieval, we drew a hydrophobic barrier for the sections. Once the barrier was dry, we added ~ 5 drops of RNAscope Protease III (ACDBio, Cat# 322337) and put the slides in the hybridization oven at 40°C for 30 min. During this incubation, we put

the probes in a 40°C incubator for 10 min and then at room temperature at least for 10 min. Following the protease incubation, we washed the slides in distilled water twice and then added the probes for 2 h in the hybridization oven at 40°C. After probe incubation, we added Amp 1 and 2 solutions successively for 30 min in the hybridization oven at 40°C and then Amp 3 solution for 15 min. After each step the slides were washed twice with the ACD wash buffer for 2 min each. We then developed the HRP-C1, HRP-C2 and HRP-C3 signals successively in the hybridization oven at 40°C, followed by a final 15 min incubation with HRP blocker. Each of these steps was followed by 2 washes for 2 min with ACD wash buffer. Finally, slides were DAPI-stained and mounted with 1-2 drops of ProLong Gold Antifade Mountant. Slides were imaged using a FV3000 confocal microscope (Olympus) with 20x objective and 60x oil-immersion objective (1.3 oil). Visualization and background removal (rolling ball radius) were done using ImageJ. Pseudo-colors were used for better visualization.

Gene panel and tissue sample preparations for MERFISH

Multiplexed RNA *in situ* hybridization was performed using MERFISH technology³⁰ and the commercialized platform, MERSCOPE (Vizgen, MA, US). A 300 gene gene panel was designed with the Vizgen Gene Panel Portal (<https://portal.vizgen.com/>) and manufactured by Vizgen. An overview of the targeted genes was shown in Fig. 3B and the detailed gene codes were provided by Table S8. 10 µm (E15.5) or 7 µm (E9.5) thick sagittal sections were prepared as described above for RNAscope. Tissue sections were mounted onto MERSCOPE Slides (Vizgen), and incubated for 20 minutes at -20°C. Next, slides were equilibrated to room temperature for 2 minutes and then immediately fixed with 4% paraformaldehyde for 15 minutes at room temperature. Slides were washed with phosphate buffered saline and then stored in 70% ethyl alcohol for up to three weeks at 4°C. MERFISH was performed according to manufacturer's instructions for fresh and fixed frozen tissue with minor modifications (Vizgen). In brief, tissue was quenched with a MERSCOPE Photobleacher for three hours and stained with the Cell Boundary Staining Kit (Vizgen). All steps were performed without parafilm covering to avoid tissue disruption. After 48 hours of probe incubation at 37°C, slides were embedded and cleared using the manufacturer's guidelines for resistant fresh frozen or fixed frozen tissue. Imaging was performed on the MERSCOPE Platform using a 300 gene imaging kit.

MERFISH data processing

Captured raw images were processed with the MERSCOPE Instrument Software. The resulting output data included a list of all detected transcripts and their spatial coordinates (CSV), mosaic images (TIFF), experiment metadata (JSON), and cell segmentation data for MERSCOPE Visualizer software (Vizgen). CellPose⁵⁴ was used for segmentation. E9.5 segmentation was based on nuclear staining, whereas cell boundary staining was used at E15.5. The output files were further analyzed using the Scanpy⁵⁵, Squidpy⁵⁶, and Seurat⁵⁰ packages. Data normalization and batch effect removal was performed using Scanpy. These criteria were used for cell and gene filtering: 1) cell, >1 gene, cell number > 3, 50 < cell volume < 2000; 2) gene, counts > 20. All subsequent analysis was performed on ECs and SMCs only. For *in silico* EC selection, we calculated the EC score for each cell using the Scanpy's score_genes function based on these marker genes: *Etv2*, *Kdr*, *Pecam1*, *Cdh5*, then positive cells (>0.8) were isolated from the whole dataset. The same method was used for SMC selection with *Myh11* and *Acta2* serving as marker

genes and selection threshold >2. Functions of Scanpy, pp.pca and pp.neighbors, were used to do principal component analysis and compute a neighborhood graph of the observations, respectively. Then Leiden clustering was performed to cluster cells according to their gene expression profile and visualized using UMAP. Scanorama³¹ was used for dataset integration, including the integration of MERFISH and scRNA-seq datasets. Gene expression values represent log-transformed, normalized expression values.

Bulk RNA sequencing

To clarify the effects of tamoxifen induction on endothelial gene expression we performed bulk RNA sequencing with purified ECs. Cdh5(PAC)-CreERT2; R26^{fsTRAP} mice were used. Tamoxifen induction was performed at E12.5 as described above. Embryos were harvested from treated (Tam⁺) and untreated (Tam⁻) mice at E15.5. Both groups had 4 replicates and 3 embryos were pooled together for each replicate. These embryos were dissociated into single-cell suspensions using Collagenase Type II (Worthington, Cat# LS004174) and 50% (v/v) Accutase (StemCell Technologies, Cat# 07920) and washed with PBS supplemented with 0.5% bovine serum albumin (BSA). The GFP⁺ ECs and GFP⁻ non-ECs were sorted using a Sony SH800 Cell Sorter (Sony, Tokyo, Japan). Then total RNAs were extracted with TRIzol reagent (ThermoFisher, Cat# 15596018) and purified with RNA Clean & Concentrator-5 Kit (ZYMO, Cat# R1013). RNA-seq libraries were prepared under the instruction of the Protocol for use with NEBNext rRNA Depletion Kit v2 (New England Biolabs, manual E7770_E7775). NEBNext rRNA Depletion Kit v2 (New England Biolabs # E7405L) and NEBNext Ultra II RNA Library Prep with Sample Purification Beads (New England Biolabs # E7775S) were used. Library quality control and next generation sequencing were performed by MedGenome Inc. (USA, DE).

RNA-seq analysis

RNA-seq mapping and quantitation was done using STAR v. 2.6.1 with flags --quantMode TranscriptomeSAM --outSAMstrandField intronMotif with --genomeDir pointing to a mm10 STAR index. The mapped reads were further analyzed by HTSeq-count v.0.11.2 and annotated using a RefSeq database. Reads count were normalized and compared by DEseq2 v.1.32.2.

Quantification and Statistical Analysis

Unless otherwise stated in the figure legends, data were expressed as mean ± SD. After testing for normality with the Shapiro-Wilke test, means were compared using unpaired two-tailed Student's t-tests (two groups) or analysis of variance (ANOVA) (3 or more groups). No exclusion criteria were applied for all analyses. Statistical analyses were performed using GraphPad Prism v.9 software (GraphPad Software Inc.) or R. P < 0.05 was considered statistically significant.

Data Access

The data have been deposited to GEO: MERFISH: GSE247450; Bulk RNAseq: GSE247449; scRNA-seq: GSE216970. Raw MERFISH data are available at <https://zenodo.org/records/10655724>.

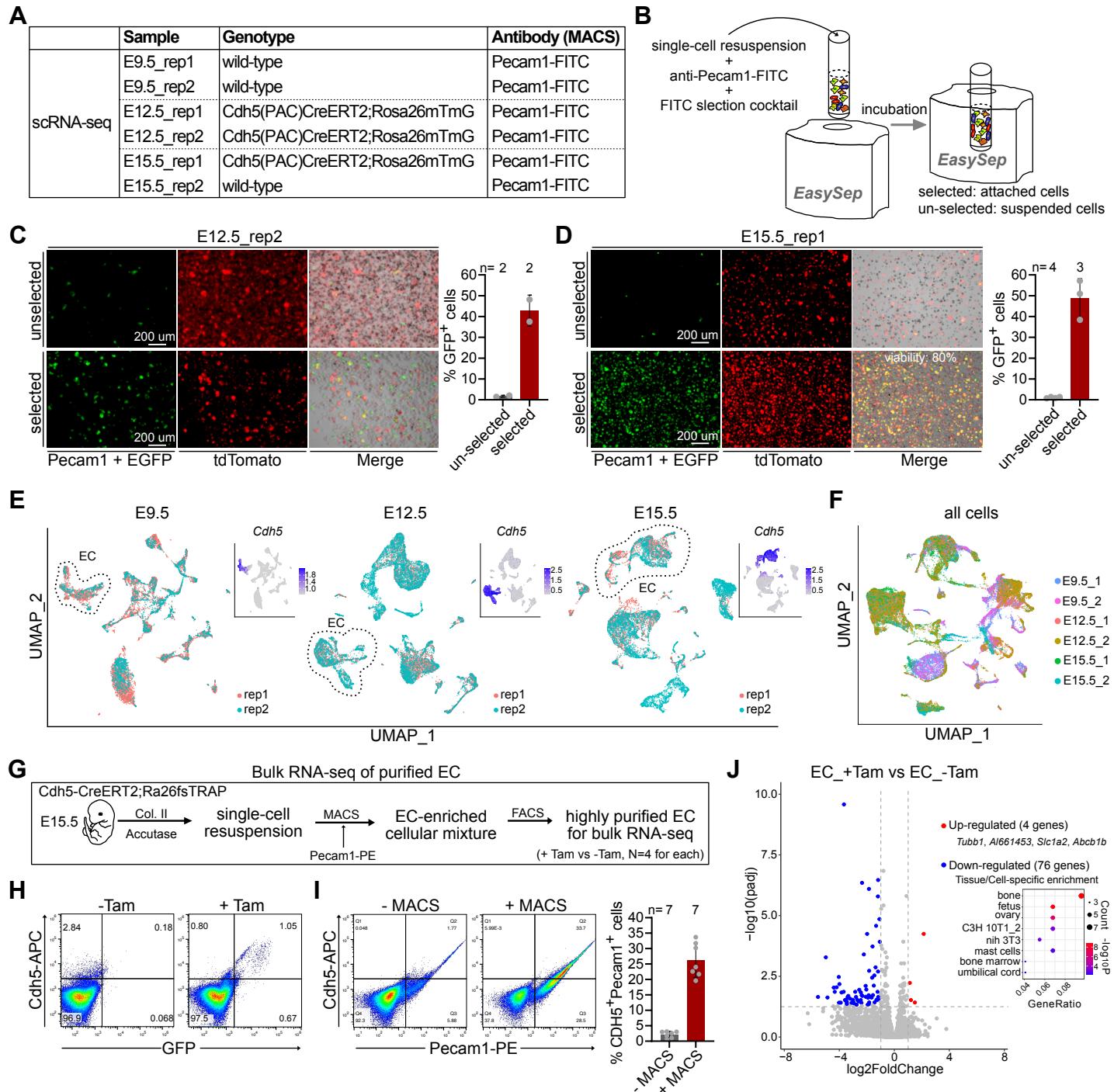


Figure S1. Minimal batch effects of scRNA-seq samples. (A) Genotype and antibody information about scRNA-seq samples. (B) Scheme of a high viability EC enrichment method used for scRNA-seq experiments. (C, D) Representative pictures and quantitative data showing the enrichment efficiencies for E12.5 and E15.5 samples. (E) UMAP plots showing minimal batch effects between replicates from the same time points. (F) UMAP plot showing minimal batch effects among all samples from E9.5, E12.5 and E15.5. (G) Scheme of bulk RNA-seq experiments. Tam: tamoxifen. (H, I) Representative flow assay data showing the Tamoxifen induction and MACS method significantly enriched ECs from embryos. (J) Differential Gene Expression (DGE) and Gene Ontology (GO) enrichment analyses of bulk RNA-seq data showing that tamoxifen induction had no significant effect on endothelial gene expression. n=3 embryos/group. Thresholds for DGE: Padj < 0.05, | log₂fold-change | > 1.

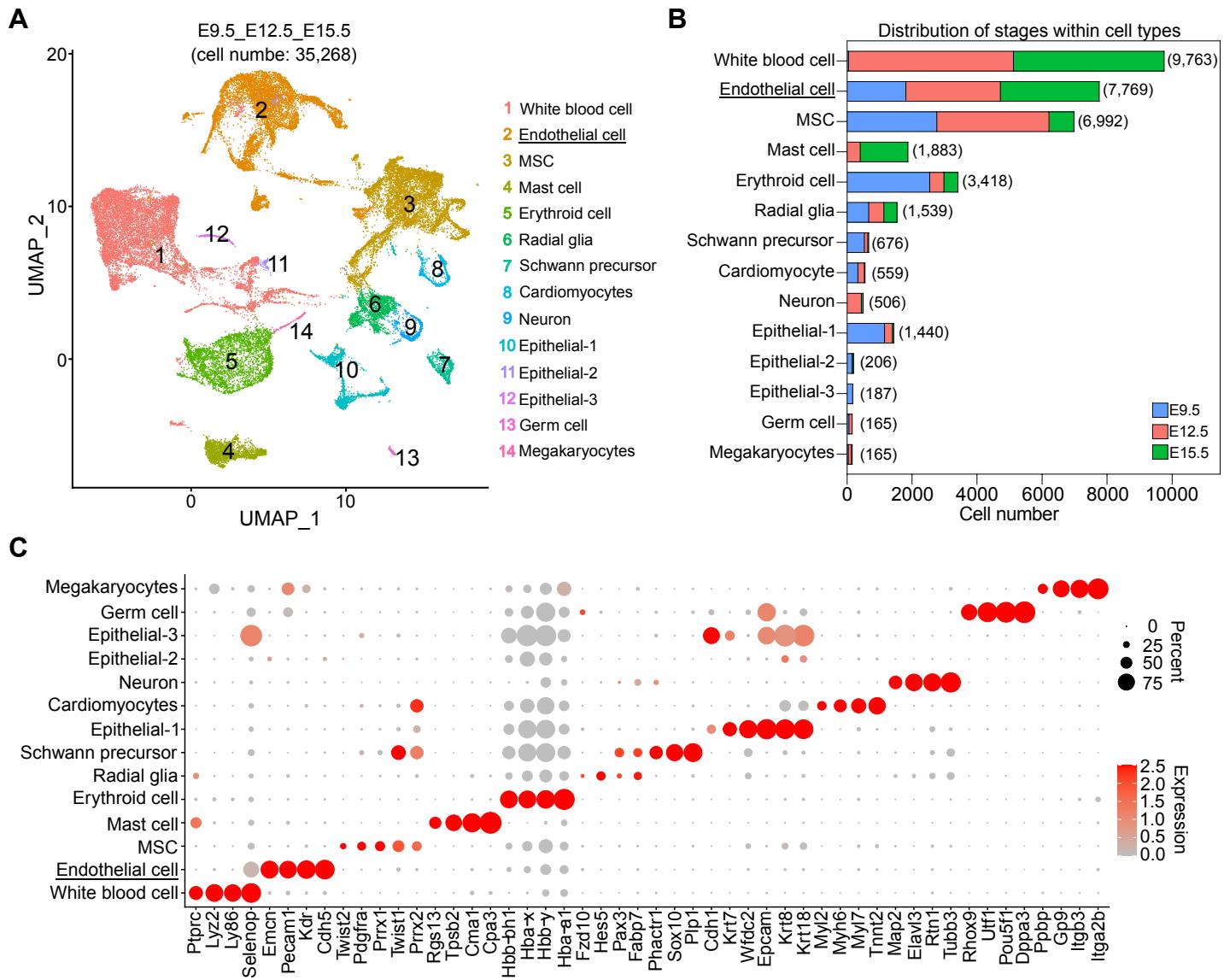


Figure S2. Identification of major cell types within scRNA-seq samples. (A) UMAP plot of the 35,268 cells colored by their identities. MSC, mesenchymal stem cells. (B) Distribution of each cell type across different embryonic stages. Total cell number for each cell type is indicated in parentheses. (C) Dot plot of key marker genes used to identify each cell type. Dot size indicates percentage of cells expressing marker genes within the cluster, and the color indicates average marker expression.

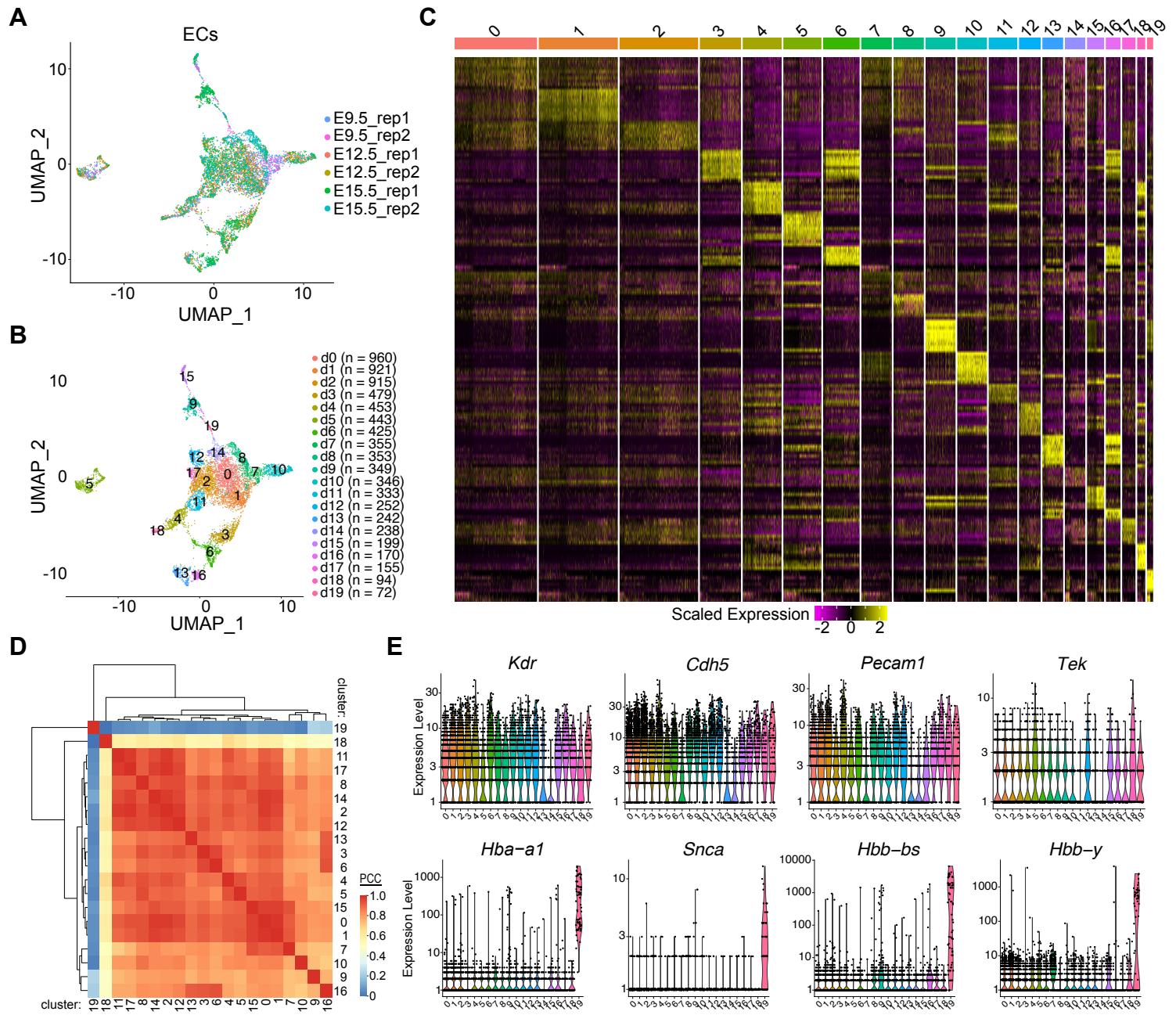


Figure S3. Embryonic EC clusters and exclusion of a red blood cell cluster. (A) UMAP plot showing stage and replicate origins of all *in silico*-selected ECs. (B) UMAP clustering of all *in silico*-selected ECs. Twenty clusters were identified. (C) Gene-expression heatmap of the top 10 marker genes for each cluster. Marker genes were identified by Seurat's FindMarkers. (D) Heatmap of the pairwise Pearson correlation coefficient (PCC) between clusters. Cluster 19 had the lowest correlation with the other clusters. The analysis was based on total enriched signatures of each cluster identified by the Seurat FindAllMarkers function. (E) Violin plots of the expression of established endothelial markers (top row) and red blood cell markers (bottom row) generated by default Seurat VlnPlot() function.

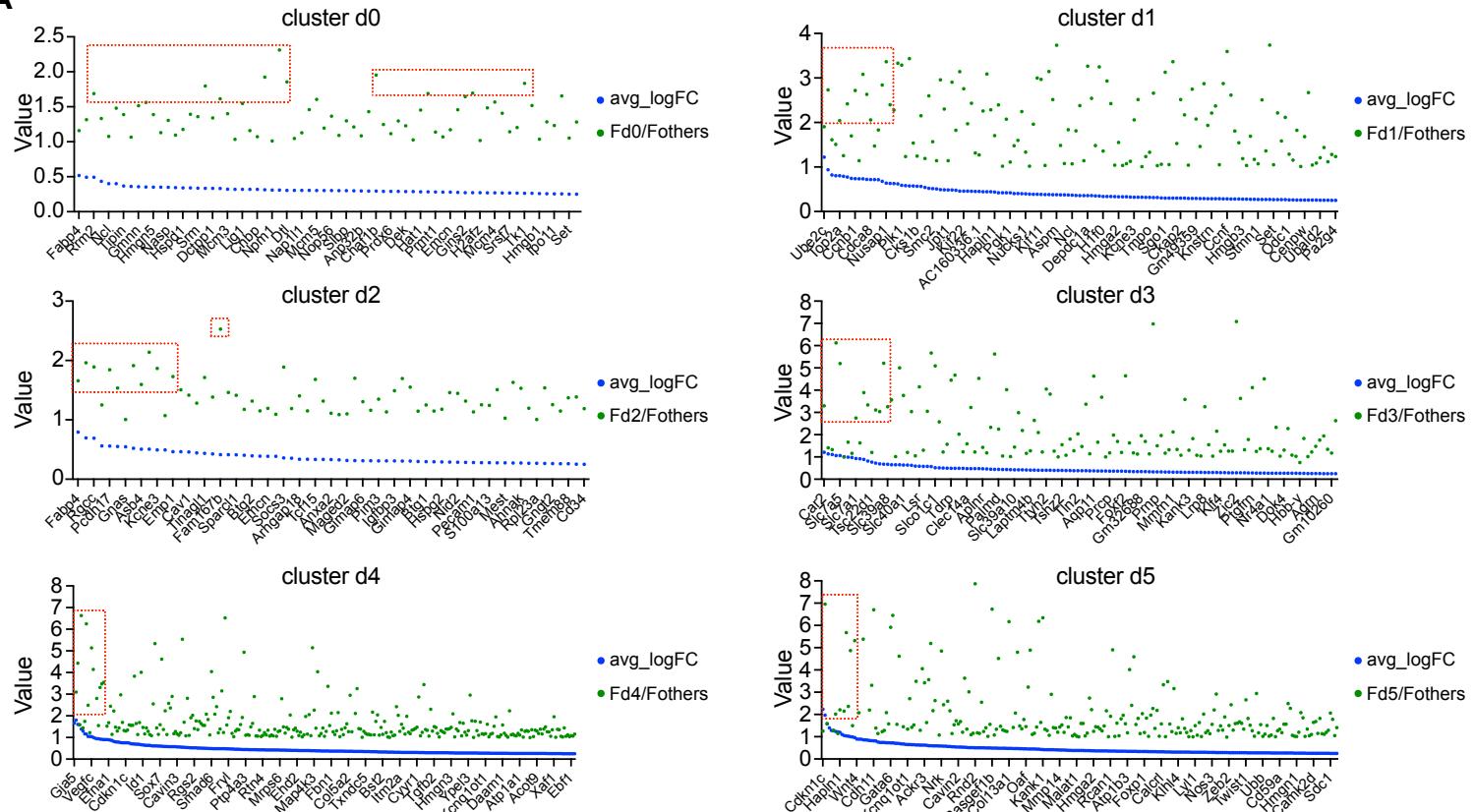
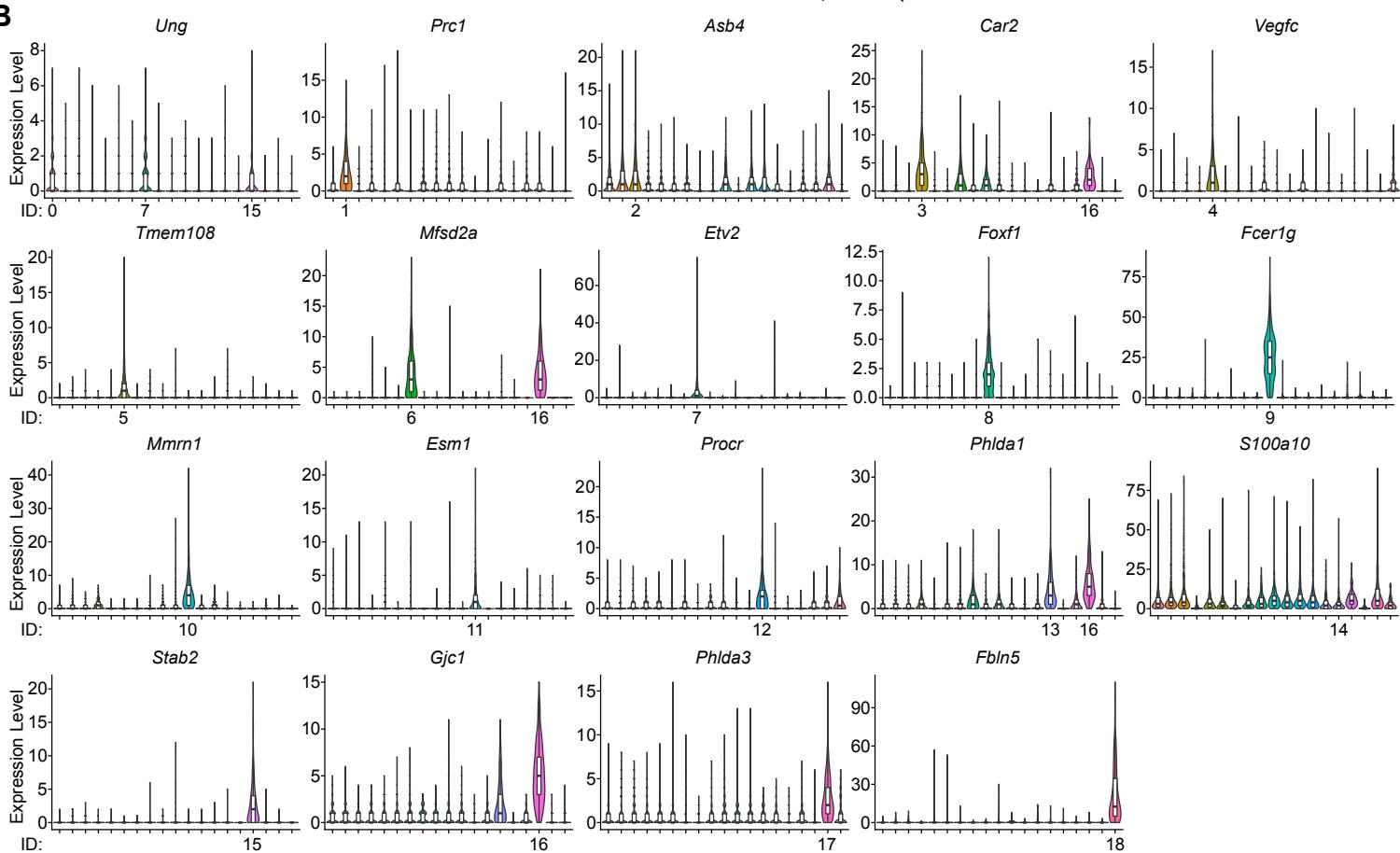
A**B**

Figure S4. Selection of specific marker genes for EC clusters. (A) Representative plots illustrating the strategy used to identify the top 10 markers for embryonic EC clusters. avg_logFC denotes natural log fold-change of the average expression between the indicated group and other groups. Positive values indicate that the gene is more highly expressed in the first cluster. F is the fraction of cells expressing the gene. High value of (FclusterN/Fothers) indicates that the gene is more widely expressed in the cluster N. Red box denotes the top 10 markers selected based on these 2 parameters. (B) Violin plots of the expression of marker genes selected through the method shown in (A). Default Seurat VlnPlot() function was used to produce these plots.

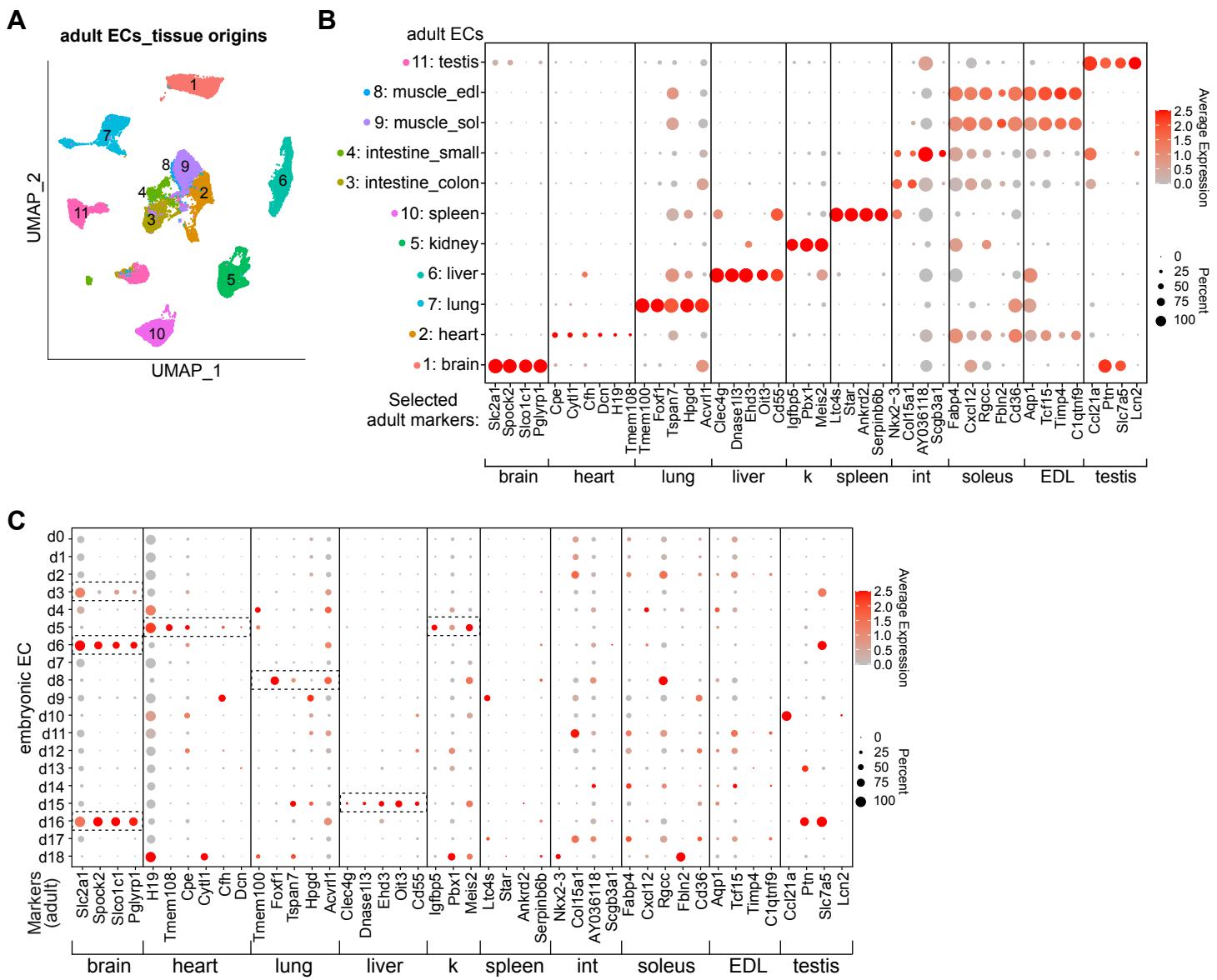


Figure S5. Determination of tissue-specific heterogeneity of embryonic ECs by marker genes. (A) UMAP plot of adult endothelial cells (published dataset E-MTAB-8077 from Kalucka et al) showing their tissue origins. (B) Dot plot showing the expression pattern of selected adult tissue-specific endothelial markers within adult ECs. k: kidney. int: intestines. (C) Dot plot showing the expression pattern of selected adult tissue-specific endothelial markers within embryonic ECs. These dot plots were generated by default Seurat DotPlot function.

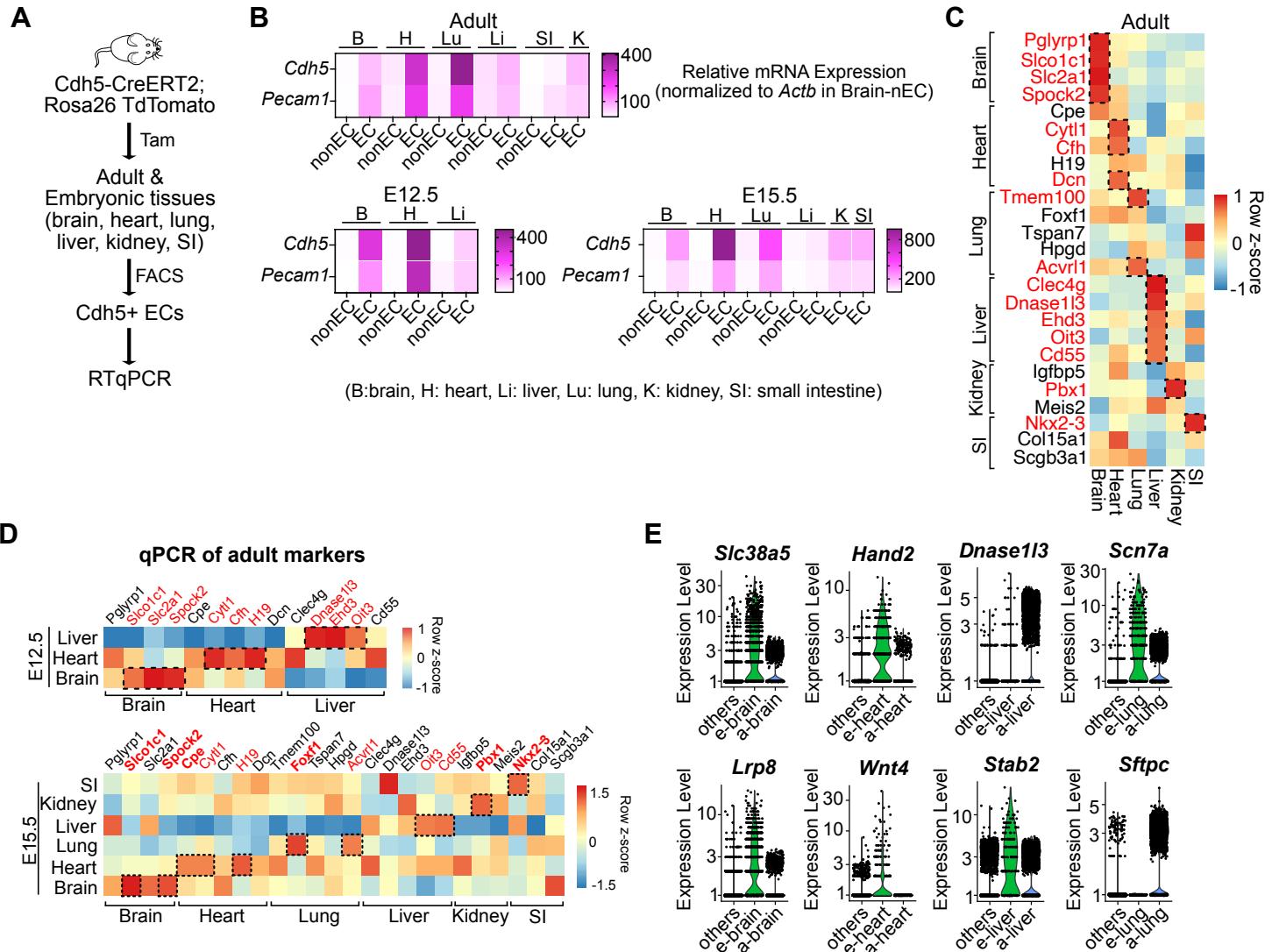


Figure S6. Identification of embryonic tissue-specific EC marker genes with FACS-qPCR and scRNA-seq. (A) Schematic illustration of experiment design for marker genes verification with FACS-qPCR. (B) Heatmap of pan endothelial markers expression in sorted ECs and nonECs, indicating success of cell sorting. (C) Heatmap of selected adult tissue-specific markers expression in purified adult tissue-specific ECs. Red genes show high specificities in qPCR experiments. SI: small intestine. (D) Heatmap of adult tissue-specific marker gene expression in sorted embryonic ECs detected by qPCR. Red and bold genes show high specificities in both qPCR and MERFISH experiments. (E) Violin plot showing developmental-stage and tissue-specific markers for ECs. e: embryonic. a: adult. Others: other all endothelial clusters except the denoted tissue clusters.

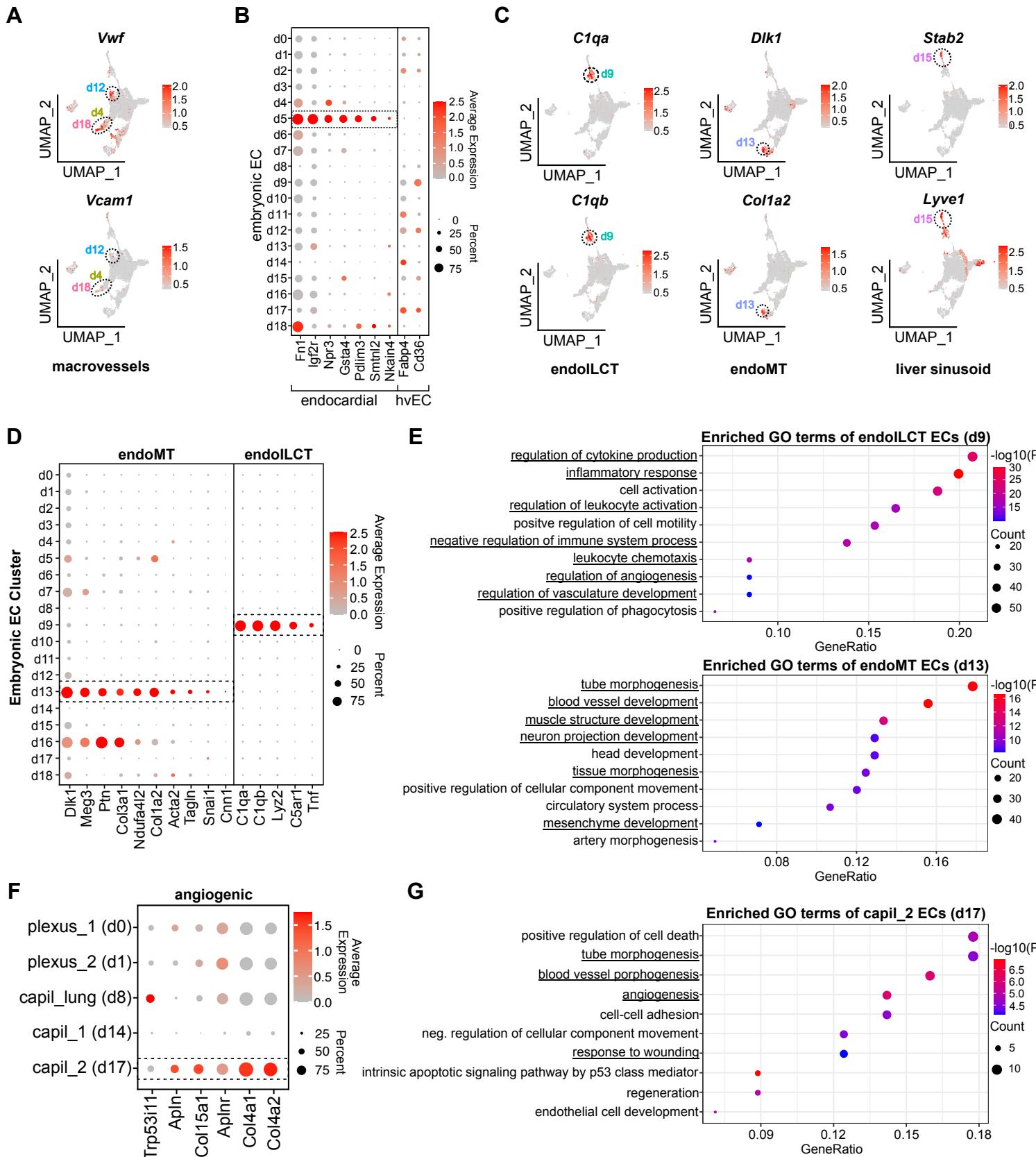


Figure S7. Identification of specialized endothelial clusters. (A) UMAP plot showing expressions of macrovessel markers. (B) Dot plot showing expressions of known endocardial and heart vascular EC markers within embryonic EC clusters. hvEC: heart vascular ECs. (C) UMAP plot showing representative markers expression identifying denoted subtypes, endoILCT (Endothelial-to-Immune-like Cell Transition, d9), endoMT (Endothelial-Mesenchymal Transition, d13) and liver sinusoid EC (d15). (D) Dot plot showing expressions of established markers of endoMT and endoILCT in embryonic EC clusters. (E) Dot plot showing the top 10 enriched GO biological process terms of cluster d9 (upper) and d13 (lower). The size of the dot indicates the number of genes annotated to each term, and the color denotes the statistical significance of the enrichment. (F) Dot plot showing expression of known angiogenic markers. (G) Dot plot showing the top 10 enriched GO biological process terms of cluster d17, capil-2.

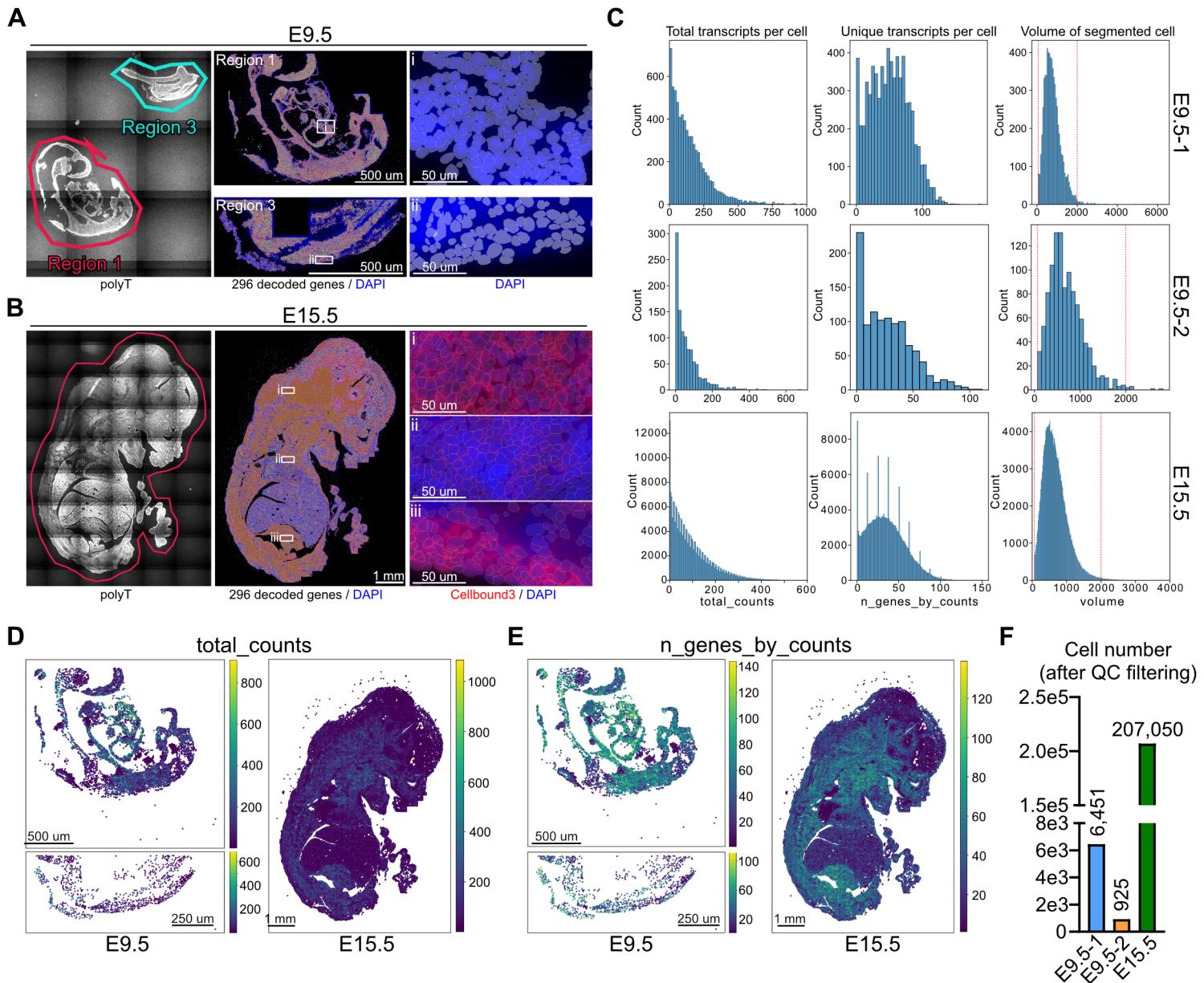
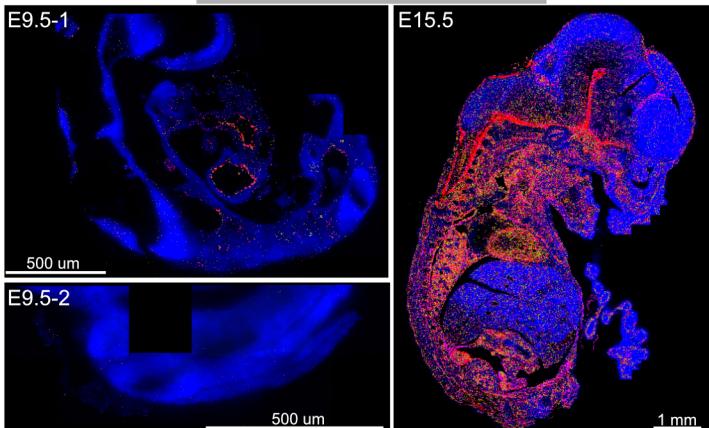


Figure S8. Quality control of MERFISH study. (A, B) Basic information about MERFISH samples obtained from E9.5 (A) and E15.5 (B) mouse embryos. Left pannels show the gross morpholgy of the section by polyT staining. Middle panels show the total signals from 296 decoded genes. Right pannels show some representative data of the cellular segmentation. All pictures were obtained by Vizgen MERSCOPE platform. (C) General information about all samples' signals . (D, E) Spatial information about the gene counts and genes distribution. (F) Bar plot showing the cell number after QC filtering from each sample.

A

Etv2/Kdr/Pecam1/Cdh5/DAPI

**B**

Myh11/Acta2/Cdh5/DAPI

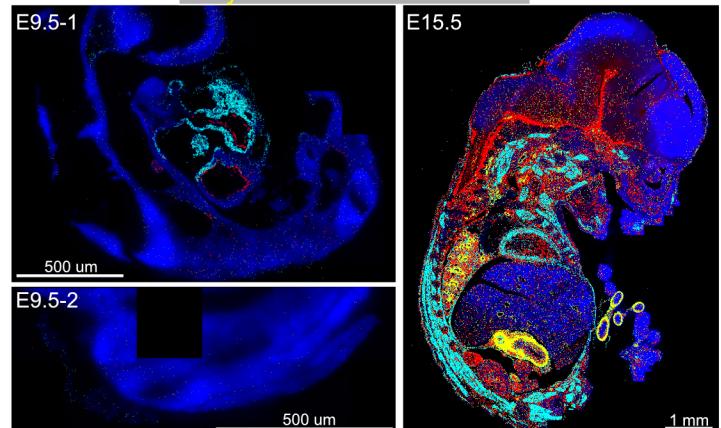
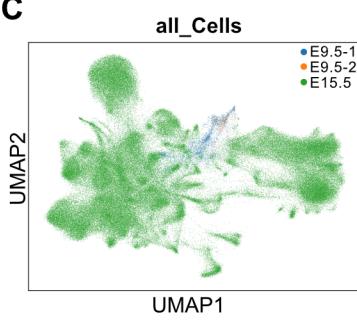
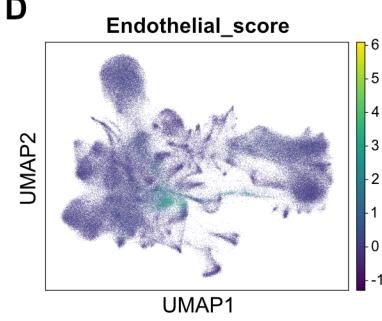
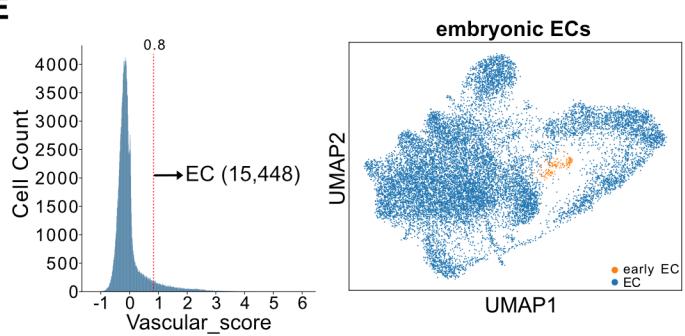
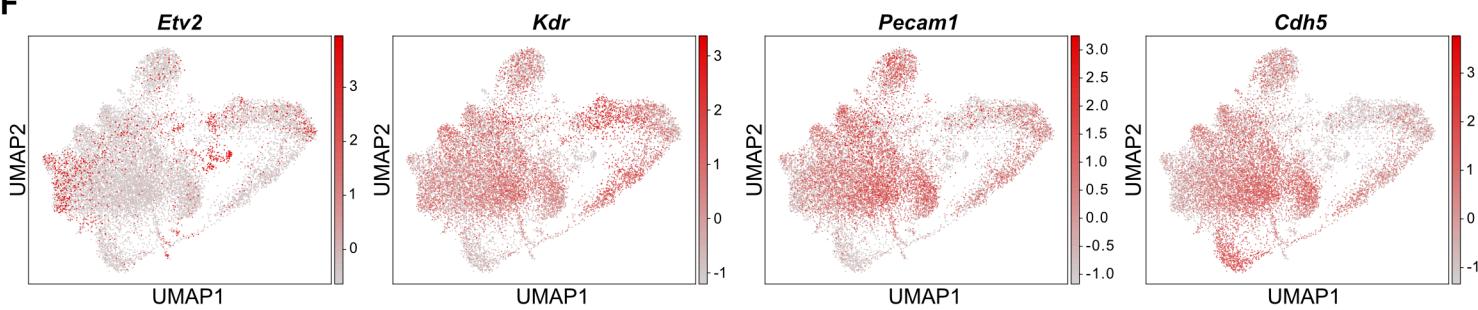
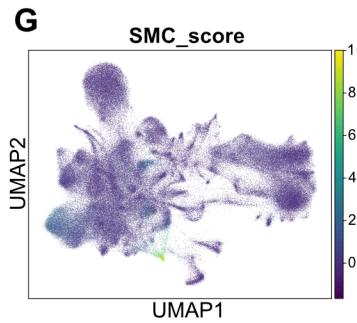
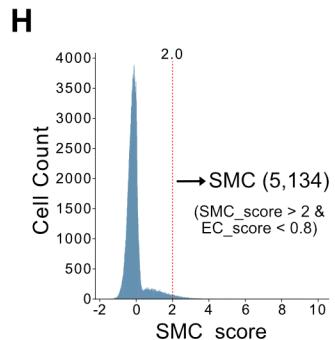
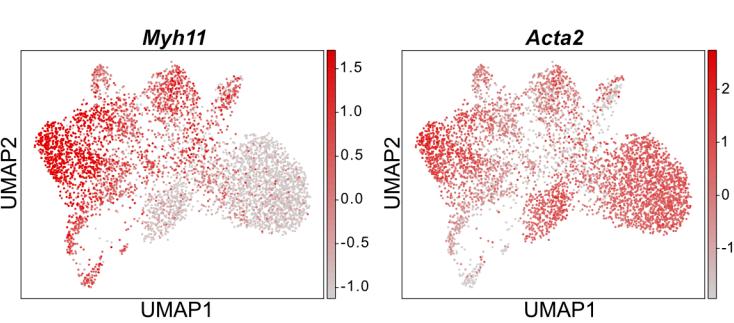
**C****D****E****F****G****H****I**

Figure S9. In silico selection of vascular cells from MERFISH data. (A, B) Microscope pictures of probe signals taken by Vizgen MERSCOPE platform. Panel A showing marker genes of earlyEC (*Etv2*) and EC (*Kdr*, *Pecam1*, *Cdh5*). Panel B showing marker genes of EC (*Cdh5*) and smooth muscle cell (SMC) (*Myh11*, *Acta2*). (C) UMAP plot of all cells sequenced by MERFISH from three samples. (D) UMAP plot showing the endothelial score for all cells. Gene expressions of *Etv2*, *Kdr*, *Pecam1* and *Cdh5* were used to calculate this score. (E) Selection of endothelial cells based on their scores (> 0.8). EC, endothelial cell. earlyEC, *Etv2*+ EC. (F) UMAP plots of marker genes expression in selected cells indicating the bioinformatic selection worked well. Color bar indicates log normalized scaled expression level of gene. (G) UMAP plot showing the SMC score for all cells. Gene expressions of *Myh11* and *Acta2* were used to calculate this score. (H) Selection of SMC based on their EC and SMC scores. (I) UMAP plots of marker genes expression in selected cells indicating the bioinformatic selection worked

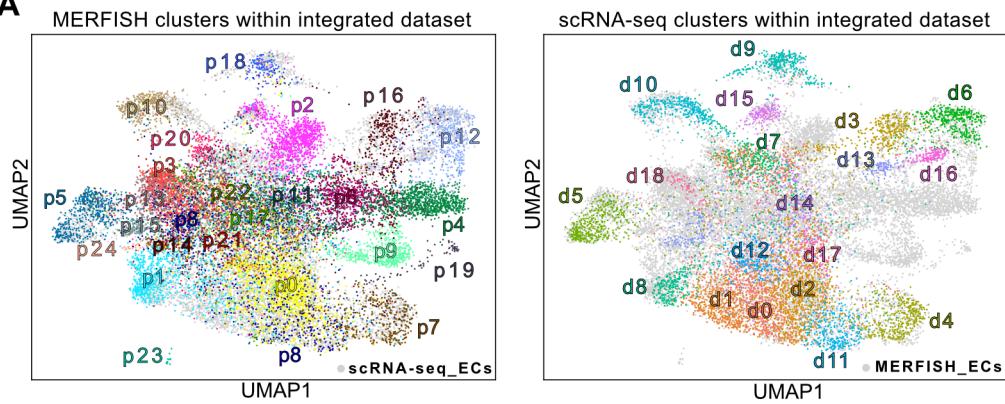
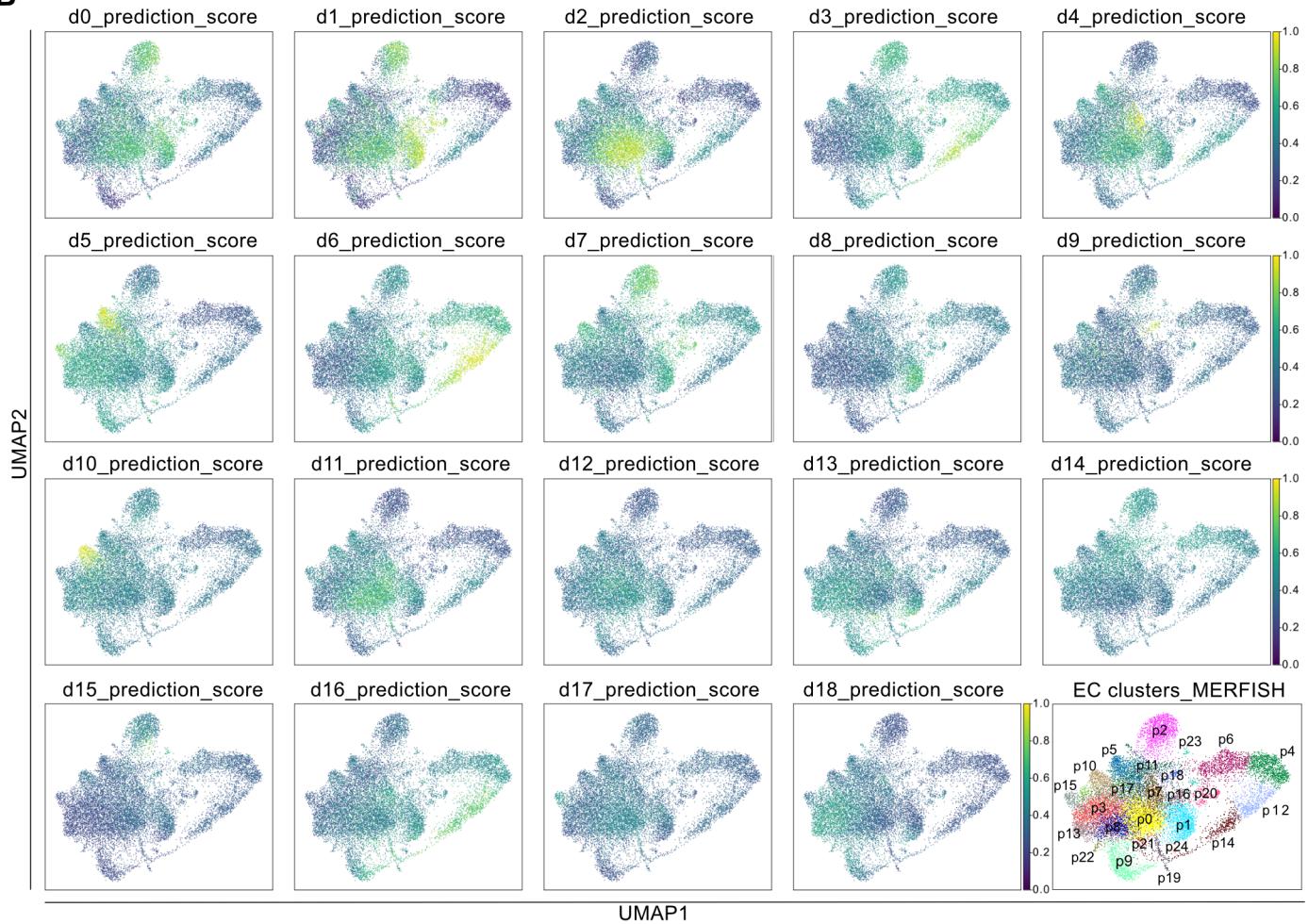
A**B**

Figure S10. Correspondence of EC clusters between MERFISH and scRNA-seq datasets.
(A) UMAP plots showing the distribution of MERFISH EC clusters (left) and scRNA-seq EC clusters (right) within integrated dataset. **(B)** UMAP plots showing the prediction score of scRNA-seq clusters measured in each cells in MERFISH dataset.

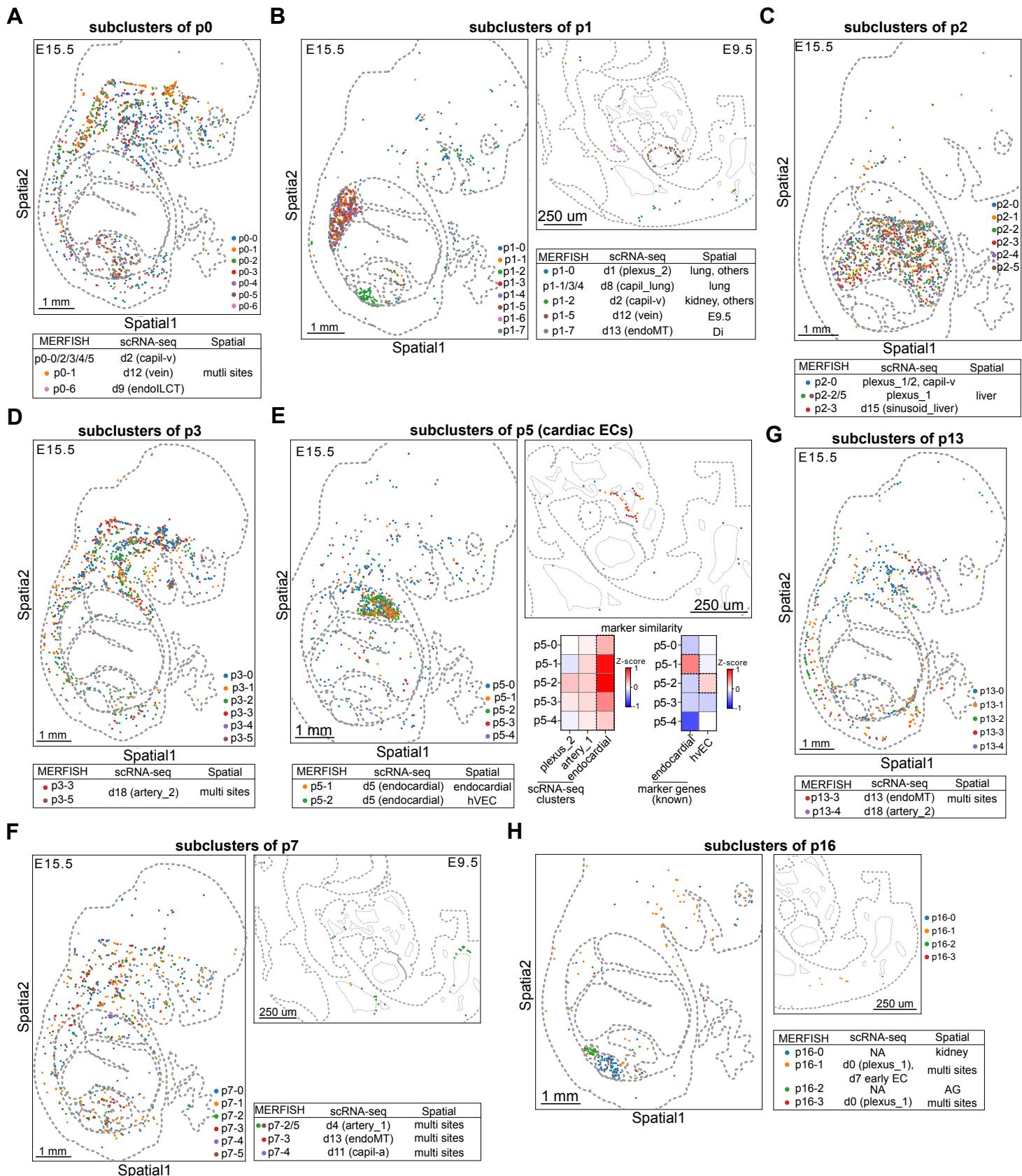


Figure S11. Spatial distribution of subpopulations of highly heterogeneous EC clusters. Spatial distribution of subclusters of MERFISH EC clusters, p0 cluster (A), p1 clusters (B), p2 cluster (C), p3 cluster (D), p5 cluster (E), p7 cluster (F), p13 cluster (G), p16 cluster (H). Tables show the correspondence clusters from MERFISH and scRNA-seq datasets. hvEC, heart vascular ECs. NA, not applicable.

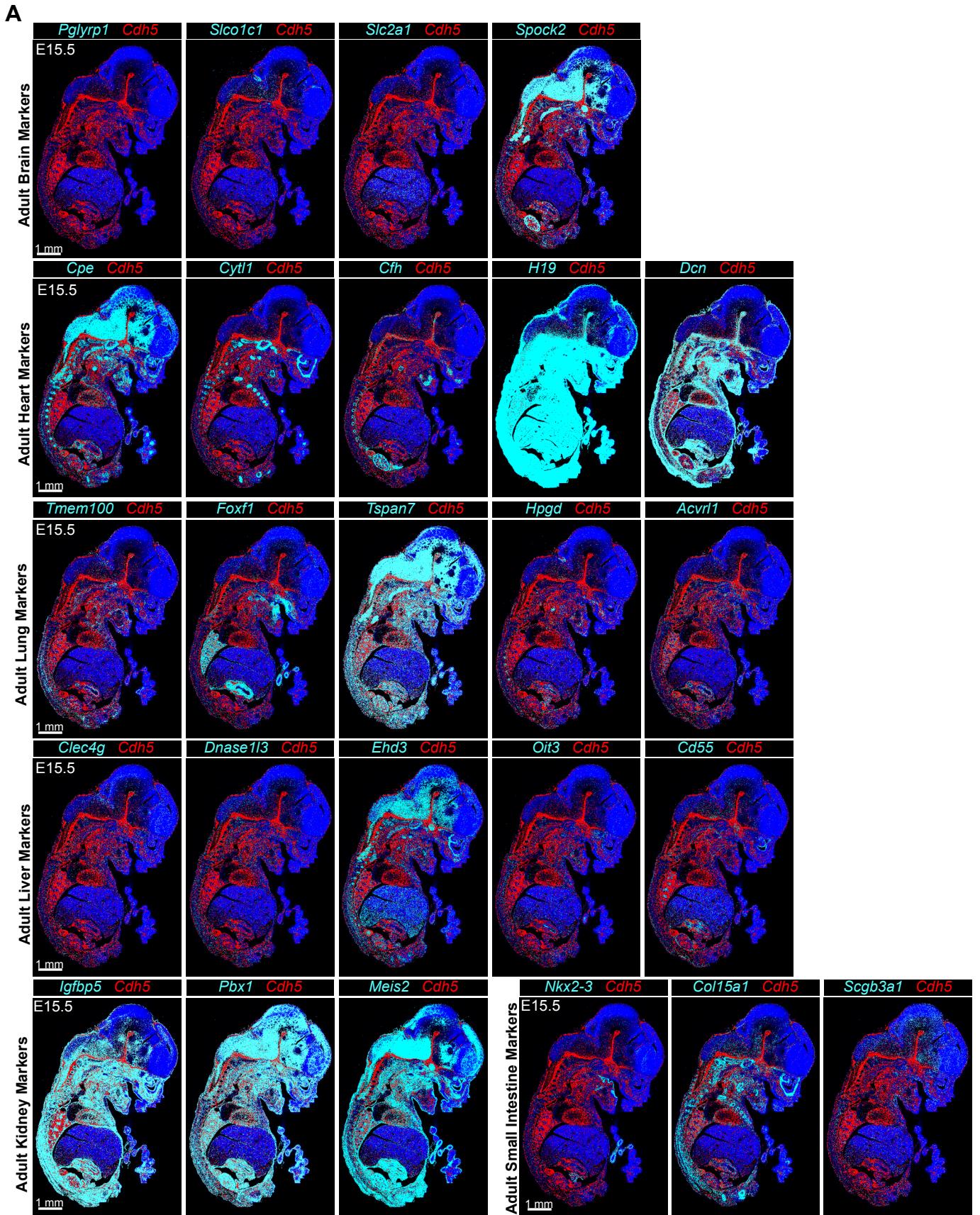


Figure S12. Spatial verification of embryonic tissue-specific EC marker genes.

(continued)

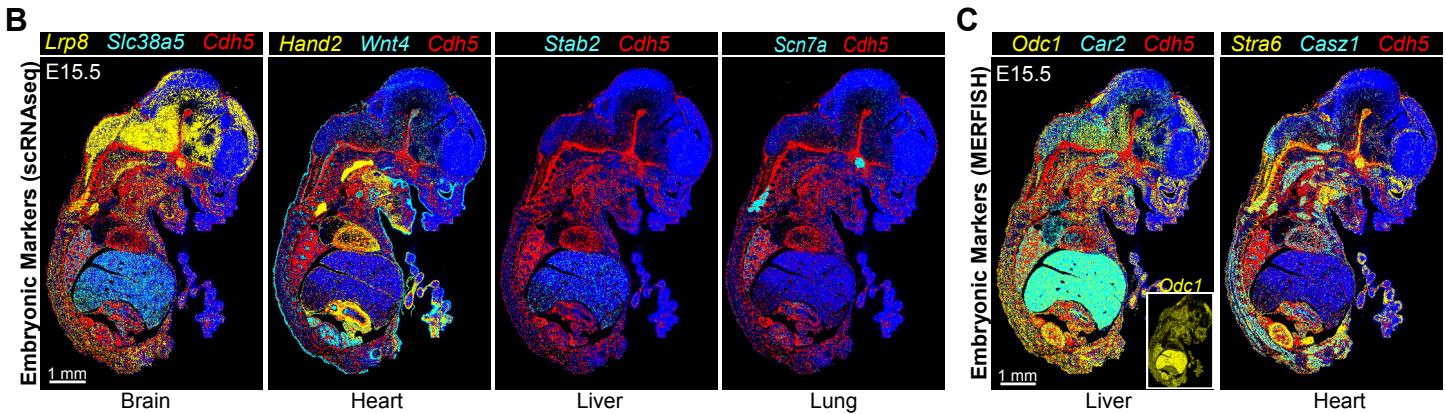


Figure S12. Spatial verification of embryonic tissue-specific EC marker genes. (A) Spatial expression of adult tissue-specific markers at E15.5. (B) Spatial expression of embryonic tissue-specific marker genes which were identified by scRNA-seq. (C) Spatial expression of embryonic tissue-specific marker genes identified by MERFISH. Blue is DAPI staining.

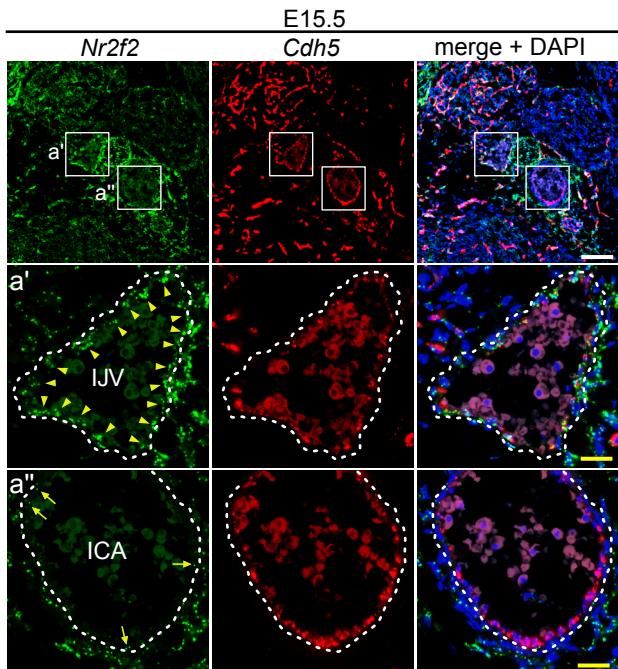
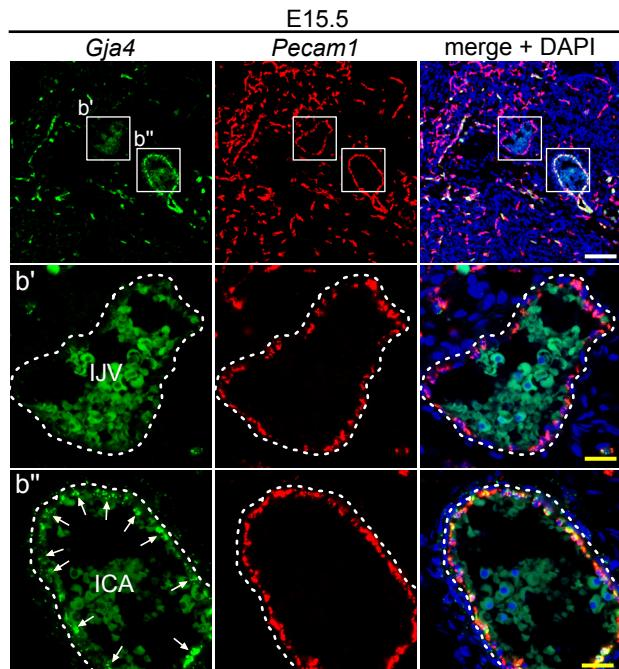
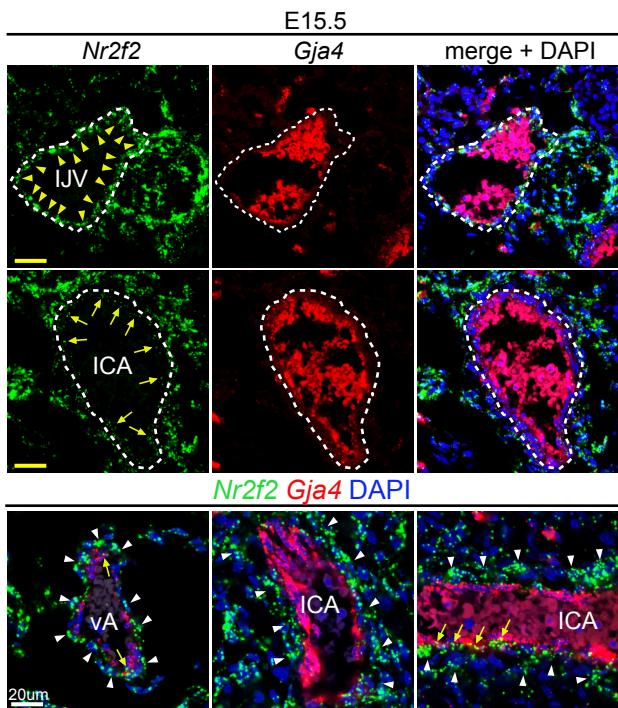
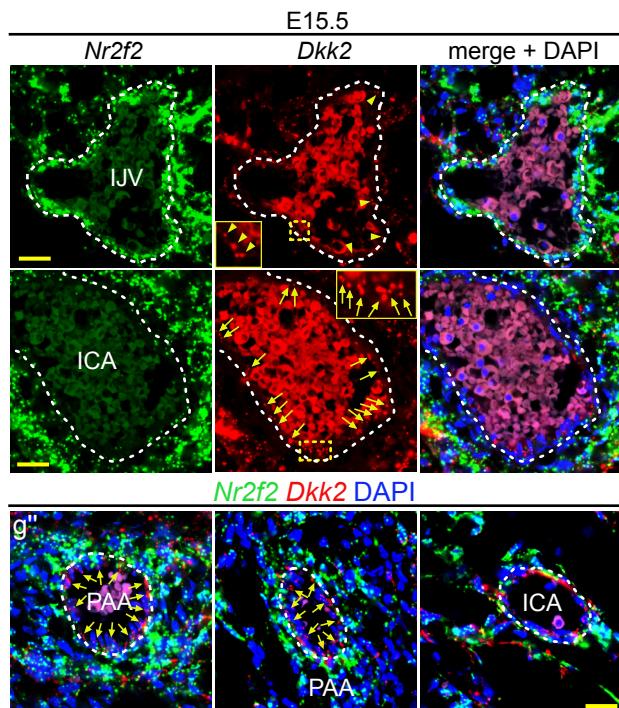
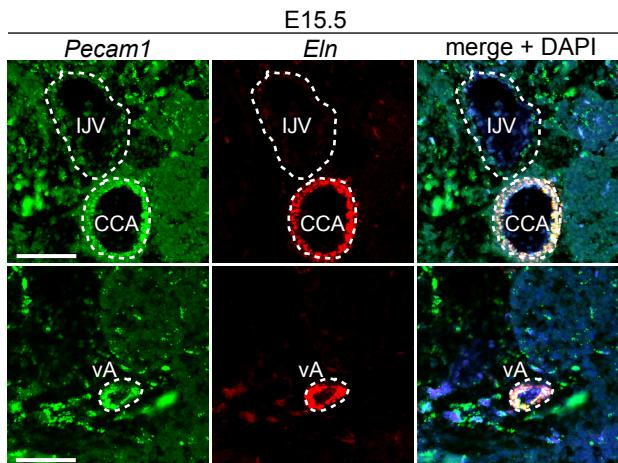
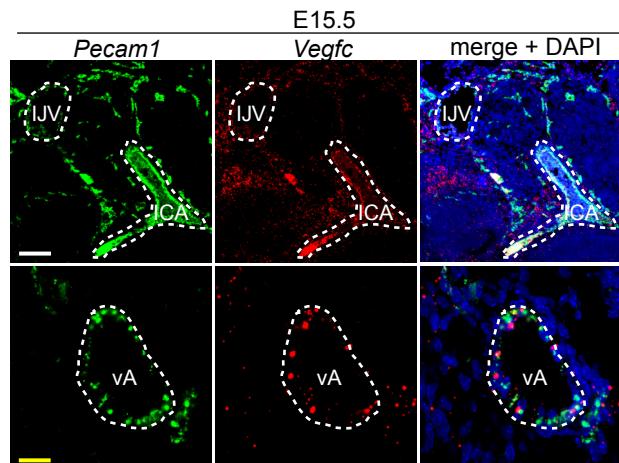
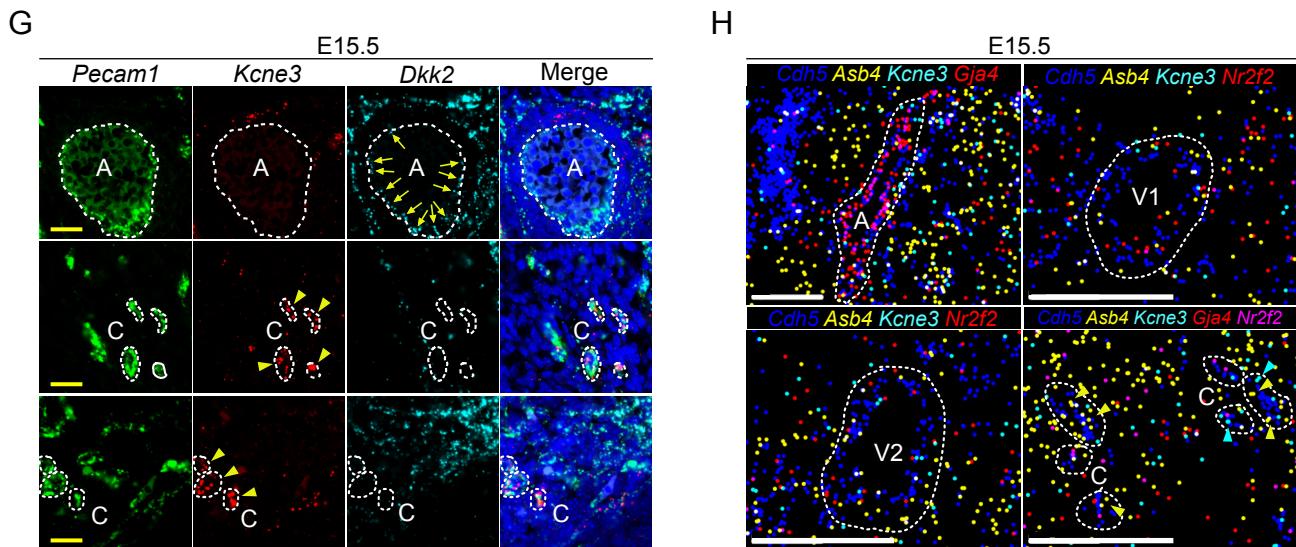
A**B****C****D****E****F**

Figure S13. Spatial expression pattern of selected marker genes. (continued)



(Figure S13, continued)

Figure S13. Spatial expression pattern of selected marker genes. (A - G) Transverse sections of E15.5 embryos stained with RNAscope probes and counterstained with DAPI. *Nr2f2* and *Cdh5* (A); *Gja4* and *Pecam1* (B); *Nr2f2* and *Gja4* (C); *Dkk2* and *Nr2f2* (D); *Pecam1* and *Eln* (E); *Pecam1* and *Vegfc* (F); *Pecam1*, *Kcne3* and *Dkk2* (G). In panels A and C, yellow arrows indicate the low expression of *Nr2f2* in aECs, yellow arrowheads indicate the high expression of *Nr2f2* in vECs and white arrowheads indicate the high expression of *Nr2f2* in non-ECs arrounding arteries. In panel B, white arrows indicate the high expression of *Gja4* in aECs. In panel D, arrows indicate the high expression of *Dkk2* in aECs and arrowheads indicate the low expression of *Dkk2* in vECs. In panel G, arrowheads indicate the high expression of *Kcne3* in capillary ECs and arrows indicate the high expression of *Dkk2* in aECs. IJV: internal jugular vein; CCA: common carotid artery; ICA: internal carotid artery; vA: vertebral artery. PAA: pharyngeal arch arteries. A: artery. C: capillary. (H) Sagital section of E15.5 embryos stained with MERFISH probes. Pictures were captured by Vizgen MERSCOPE platform and all fluorescent signals were amplified and merged by ImageJ software. Arrowheads indicate the exression of *Kcne3* and *Asb4* in capillaries. A: artery. V1: vein-1. V2: vein-2. C: capillary. Bars: White, 100 µm. Yellow, 20 µm.

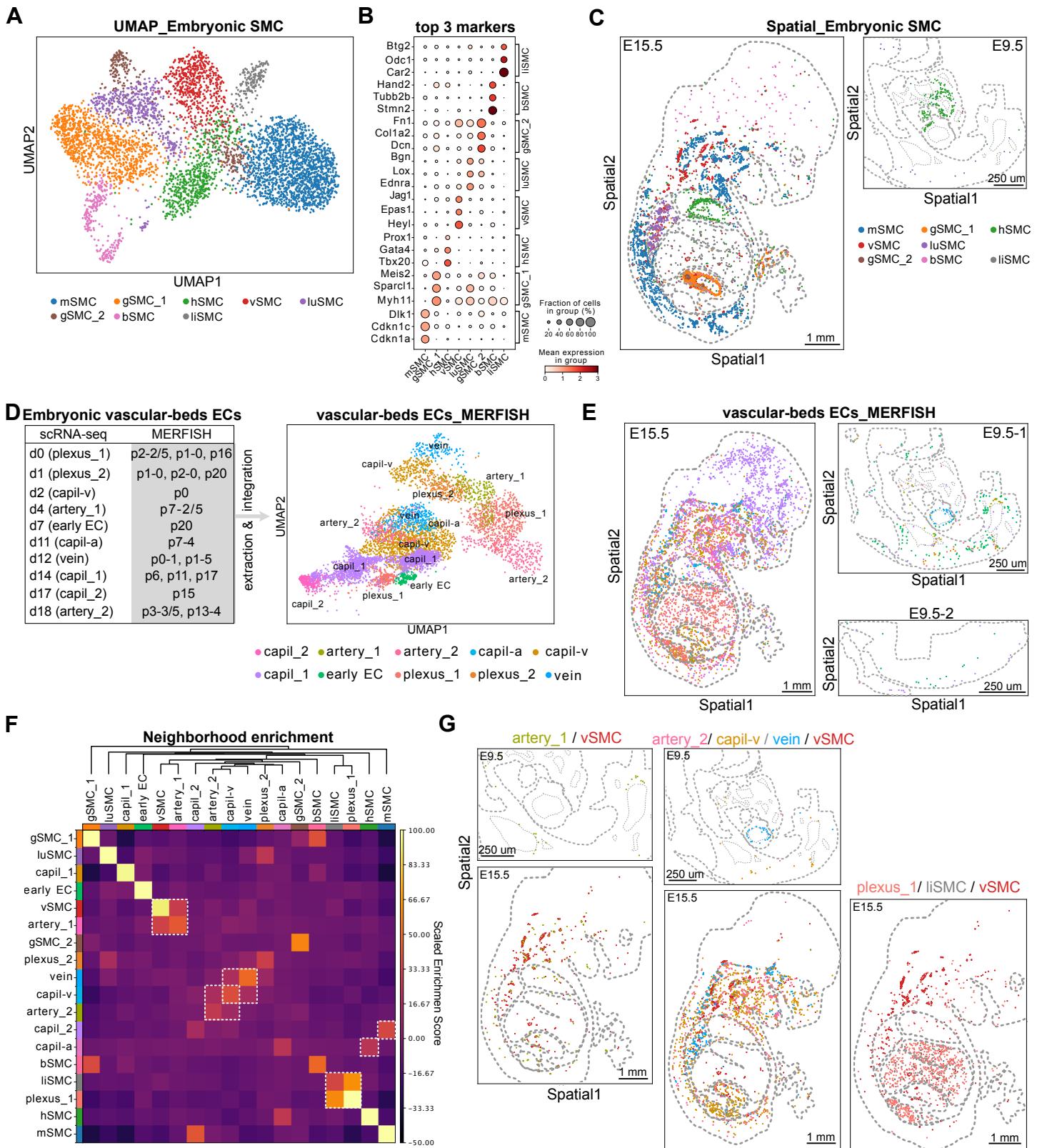
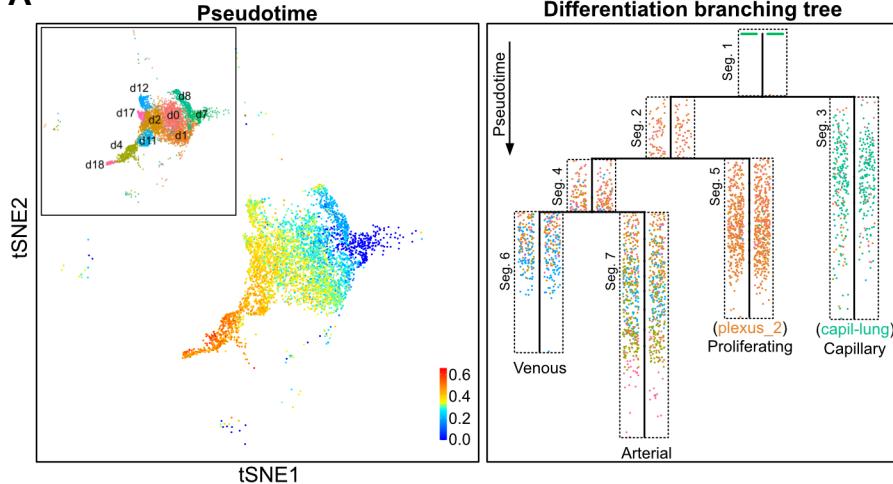
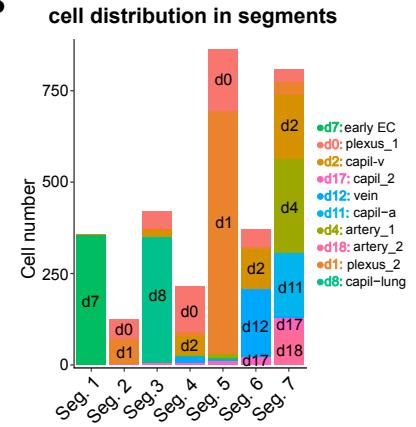


Figure S14. Spatial correlation of endothelial cells (ECs) and smooth muscle cells (SMCs). (A) UMAP of embryonic SMCs identified by MERFISH. mSMC, muscle SMC. gSMC, gastrointestinal SMC. hSMC, heart SMC. vSMC, vascular SMC. iuSMC, lung SMC. bSMC, brain SMC. liSMC, liver SMC. (B) Heatmap of top 3 markers of SMC subtypes. (C) Spatial plots showing distribution of SMCs. (D) UMAP of vascular-bed ECs identified with MERFISH. (E) Spatial plots showing distribution of vascular-bed ECs. (F) Hierarchically clustered heatmap showing EC and SMC clusters of enriched neighborhoods. (G) Spatial plots showing the distribution of clusters which have close correlations illustrated in panel F.

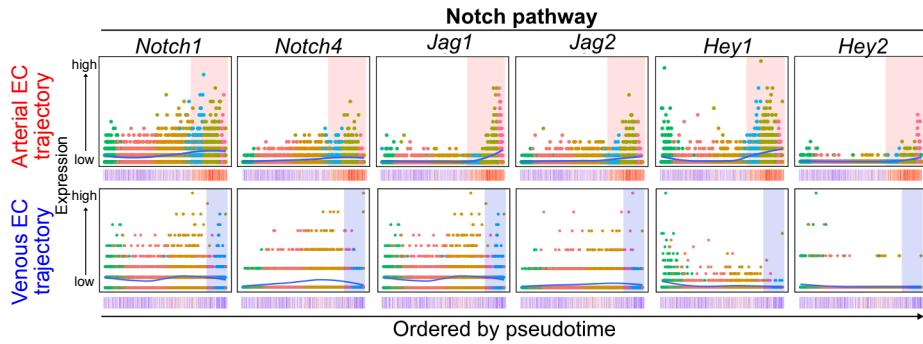
A



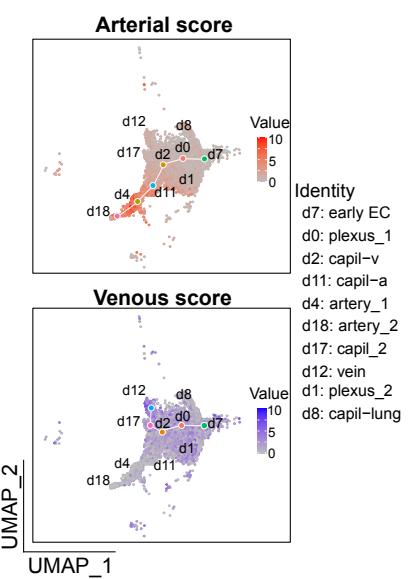
B



C



E



D

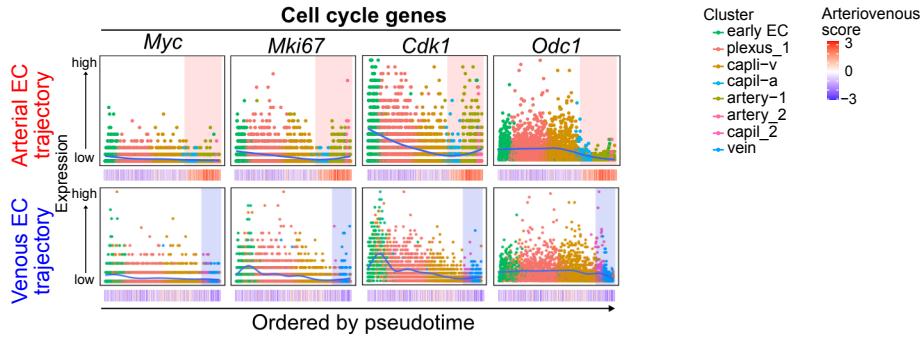


Figure S15. Characterization of arterial and venous differentiation trajectories. (A) URD-inferred path of arteriovenous differentiation. tSNE map of pseudotime (left) and differentiation paths from early EC to vEC, aEC, plexus, and capillary end points. Seg., segment. (B) Cell distribution in different segments inferred by URD. (C, D) Gene expression along the arterial and venous differentiation paths inferred by Slingshot. Notch pathway genes (B) and cell cycle genes (C). Cells are arranged along the x-axis by pseudotime. Arteriovenous scores are shown at the bottom of each plot. Red and blue shading at right of each plot indicate aECs and vECs, respectively. (E) UMAP plots showing arterial (top) and venous (bottom) scores of the individual cells in the arteriovenous EC groups.

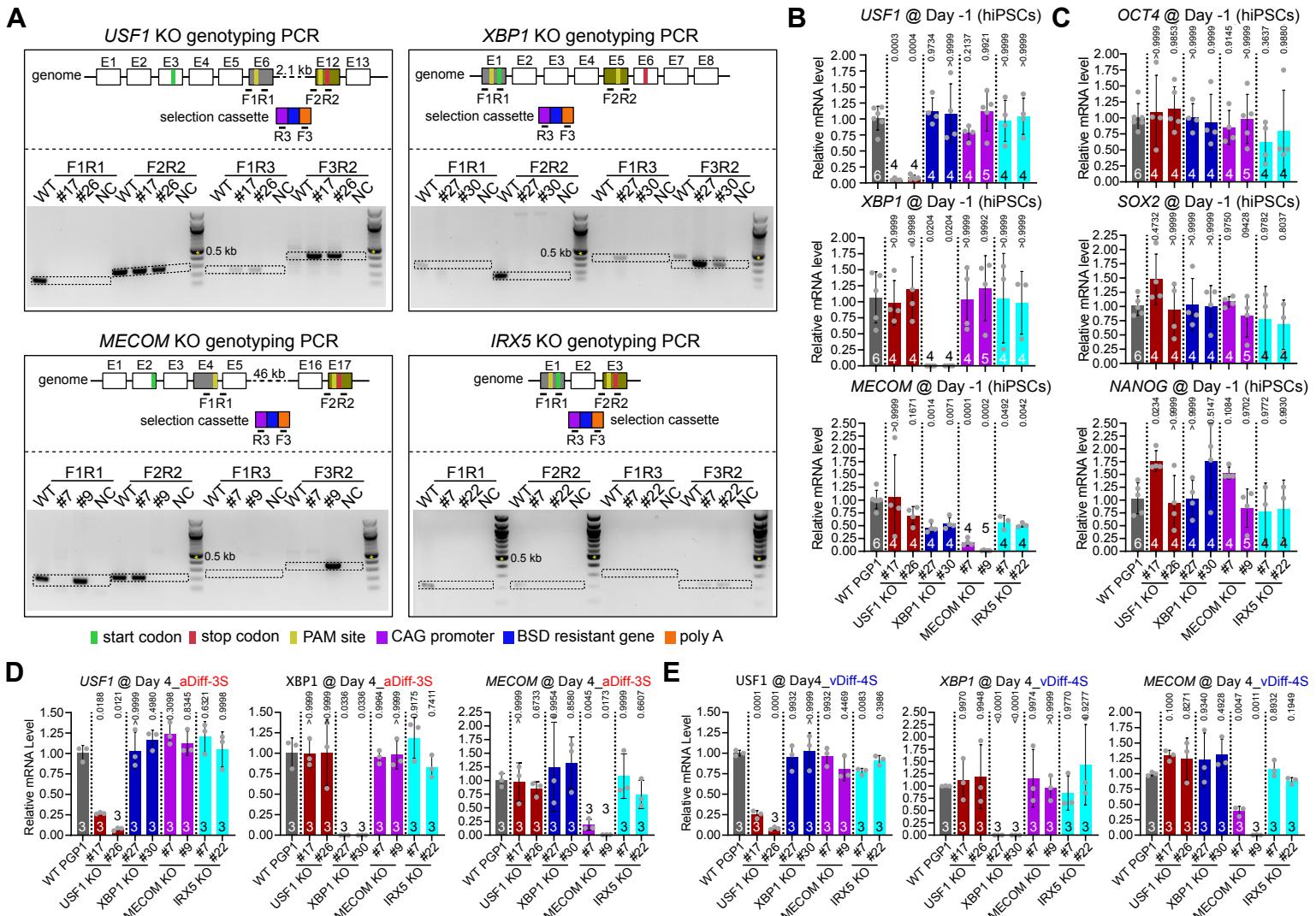


Figure S16. Knockout of selected transcription factors. (A) Construction of CRISPR knockout iPSCs. Diagrams show gene knockout and genotyping strategy for *USF1* KO, *XBP1* KO, *MECOM* KO and *IRX5* KO hiPSCs. E: exon. F: forward primer. R: reverse primer. WT: wild type. NC: negative control. 100 bp DNA ladder was used. (B, C) mRNA expression level in indicated knockout and WT hiPSC lines at Day -1, the start point of differentiation. (D, E) mRNA expression level in indicated knockout and WT lines during hiPSC-aEC (D) and hiPSC-vEC (E) differentiation. ANOVA with Dunnett's multiple comparison test vs. WT.

Table S1. Major resources table

This table summarizes the major resources used in this study.

Animals (in vivo studies)

Species	Vendor or Source	Background Strain	Sex	Persistent ID / URL
Mus musculus	Boston Children's Hospital/The Jackson Laboratory	C57BL6/J	Male and Female	

Genetically Modified Animals

	Species	Vendor or Source	Background Strain	Other Information	Persistent ID / URL
Tg(Cdh5-cre/ERT2)1Rha	Mus musculus	Taconic	C57BL6	Model #: 13073	https://www.taconic.com/mouse-model/cdh5pac-creert2-mouse
B6.Cg-Gt(ROSA)26Sor tm1(CAG-tdTomato)Hze/J	Mus musculus	The Jackson Laboratory	C57BL6/J	Strain #: 007914	https://www.jax.org/strain/007914
Gt(ROSA)26Sor tm4(CTB-tdTomato,-EGFP)Luo/J	Mus musculus	The Jackson Laboratory	C57BL6/J	Strain #: 007576	https://www.jax.org/strain/007576
B6.129S4-Gt(ROSA)26Sor tm1(CAG-EGFP/Rpl10a,-birA)Wtp/J	Mus musculus	The Jackson Laboratory	C57BL6/J	Strain #: 022367	https://www.jax.org/strain/022367

Antibodies

Target antigen	Vendor or Source	Catalog #	Working concentration	Persistent ID / URL
anti-human CD144-PE-Cy7	Thermo Fisher	25-1449-42	Flow: 1:20	https://www.thermofisher.com/antibody/product/CD144-VE-cadherin-Antibody-clone-16B1-Monoclonal
anti-human CD144-PE	Thermo Fisher	12-1449-82	Flow: 1:20	https://www.thermofisher.com/antibody/product/CD144-VE-cadherin-Antibody-clone-16B1-Monoclonal
anti-human CD144-FITC	BD Biosciences	560411	Flow: 1:20	https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/fitc-mouse-anti-human-cd144.560874
anti-human CD144-PerCP-Cy5.5	BD Biosciences	561566	Flow: 1:20	https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/percp-cy-5-5-mouse-anti-human-cd144.561566
anti-human CD31	Agilent	M082329-2	IF: 1:200	https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cd31-endothelial-cell-%28concentrate%29-76539
anti-human CD31	Abcam	Ab76533	IF: 1:200	https://www.abcam.com/cd31-antibody-epr3094-ab76533.html
anti-human CD73-BV421	BD Biosciences	562430	Flow: 1:20	https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv421-mouse-anti-human-cd73.562430
anti-human DLL4-PE	BioLegend	346506	Flow: 1:20	https://www.biologen.com/en-ie/products/pe-anti-human-delta-like-protein-4-dll4-antibody-6326
anti-human DLL4-APC	BioLegend	346508	Flow: 1:20	https://www.biologen.com/en-ie/products/apc-anti-human-delta-like-protein-4-dll4-antibody-7835
anti-human/mouse Nrp2-APC	R&D Systems	FAB22151A	Flow: 1:20	https://www.rndsystems.com/products/human-mouse-neuropilin-2-apc-conjugated-antibody-257103_fab22151a

anti-mouse CD31	Abcam	Ab56299	IF: 1:100	https://www.abcam.com/cd31-antibody-rm0032-1d12-bsa-and-azide-free-ab56299.html
anti-human/mouse CD31-FITC	Thermo Fisher	11-0311-82	Flow: 1:20 IF: 1:100	https://www.thermofisher.com/antibody/product/CD3-1-PECAM-1-Antibody-clone-390-Monoclonal/11-0311-82

Cultured Cells

Name	Vendor or Source	Sex (F, M, or unknown)	Persistent ID / URL
PGP1 hiPSC	Wang et al., 2014	Male	

Data & Code Availability

Description	Source / Repository	Persistent ID / URL
Single-cell transcriptome atlas of embryonic ECs	This study	GSE216970
Early and extensive venous arterialization during mammalian embryogenesis	Hou et al., 2022	GSE94877
Single-cell transcriptome atlas of adult ECs	Kalucka et al., 2020	ArrayExpress: E-MTAB-8077
MERFISH reveals embryonic endothelial heterogeneity	This study	GSE247450
Bulk RNA-seq reveals effects of tamoxifen on endothelial gene expression	This study	GSE247449

Other

Description	Source / Repository	Persistent ID / URL
SB505124	Tocris Bioscience	Cat# 326310
Thiazovivin	StemCell Technologies	Cat# 72252
polyvinyl alcohol	Sigma	Cat# P8136-250G
chemically defined lipid concentrate	Thermo Fisher	Cat# 11905031
1-thioglycerol	Sigma	Cat# M6145-100ML
Insulin, human recombinant	Roche	Cat# 11376497001
human transferrin	Sigma	Cat# 10652202001
penicillin/streptomycin	Thermo Fisher	Cat# 10378016
Geltrex	Thermo Fisher	Cat# A1413302
CHIR99021	StemCell Technologies	Cat# 1001042
GDC-0941	StemCell Technologies	Cat# 73152
Forskolin	Tocris Bioscience	Cat# 109910
Accutase	StemCell Technologies	Cat# 07920
Collagenase, Type 2	Worthington	Cat# LS004174
Versene	Thermo Fisher	Cat# 15040066
XAV 939	Tocris Bioscience	Cat# 374810
ascorbic acid-2-phosphate	Sigma	Cat# A8960-5G
DMH-1	Tocris Bioscience	Cat# 412610
RO4929097	Selleck	Cat# S1575
PD0325901	Tocris Bioscience	Cat# 419210
FGF-2	PeproTech	Cat# 100-18B
Recombinant Human BMP4	R&D Systems	Cat# 314-BP-050
Human Recombinant bFGF, ACF	StemCell Technologies	Cat# 78134.1
Recombinant Human VEGF 165 Protein	R&D Systems	Cat# 293-VE-050
Human Recombinant VEGF-165, ACF	StemCell Technologies	Cat# 78159
Activin A	R&D Systems	Cat# 338-AC-050
Cas9	IDT	Cat# 1074181
CD144 microbeads, human	Miltenyi Biotec	Cat# 130-097-857
EasySep™ PE Positive Selection Kit II	StemCell Technologies	Cat# 17684
EasySep™ FITC Positive Selection Kit II	StemCell Technologies	Cat# 17682
Chromium Single Cell 3' Library, Gel Bead & Multiplex Kit and Chip Kit, v2	10x Genomics	PN-120237
Chromium Single Cell A Chip Kit	10x Genomics	PN-120236
Chromium i7 Multiplex Kit	10x Genomics	PN-120262

RNAscope® Multiplex Fluorescent Reagent Kit v2	Advanced Cell Diagnostics	Cat# 323100
MERSCOPE Sample Verification Kit (Mouse)	VIZGEN	Cat# 10400008
MERSCOPE Slide Box	VIZGEN	Cat# 10500001
MERSCOPE 300 Gene Imaging Kit	VIZGEN	Cat# 10400005
MERSCOPE 300 Gene Panel	VIZGEN	Cat# 10400002
MERSCOPE Sample Prep Kit	VIZGEN	Cat# 10400012
MS Columns	Miltenyi Biotec	Cat# 130-042-201
GentleMACS C tubes	Miltenyi Biotec	Cat# 130-093-237
40 um cell strainer	Corning	Cat# 352340
GlutaMAX	Thermo Fisher	Cat# 35050061
IMDM	Thermo Fisher	Cat# 12440053
DMEM/F-12, GlutaMAX™ supplement	Thermo Fisher	Cat# 10565018
mTeSR1	StemCell Technologies	Cat# 85850
NEBNext rRNA Depletion Kit v2	NEB	Cat# E7405L
NEBNext Ultra II RNA Library Prep with Sample Purification Beads	NEB	Cat# E7775S
NEBNext Multiplex Oligos for Illumina	NEB	Cat# E6440S
TRIzol reagent	Thermo Fisher	Cat# 15596018
RNA Clean & Concentrator-5 Kit	ZYMO	Cat# R1013

Table S2: Key quality metrics of the sample libraries and sequencing process

Samples	E9.5_1	E9.5_2	E12.5_1	E12.5_2	E15.5_1	E15.5_2
Number of embryos pooled	9	10	4	4	3	3
Number of cells passed QC-filtering	7,301	5,070	6,492	9,320	5,359	7,923
Fraction Reads in Cells	93.50%	93.10%	85.40%	93.50%	92.90%	96.10%
Mean Reads per Cell	16,719	43,015	13,797	16,625	19,914	30,437
Median Genes per Cell	2,573	4,072	1,809	2,511	2,458	2,816
Total Genes Detected	19,280	19,877	20,274	21,694	20,433	21,296
Median UMI Counts per Cell	8,047	19,877	4,673	7,173	7,789	10,374
Sequencing Saturation	21.60%	32.30%	29.20%	12.10%	24.90%	36.90%
Reads Mapped Confidently to Transcriptome	72.90%	75.80%	66.10%	70.30%	73.40%	73.60%

Table S7. Oligonucleotide sequences used in this study

Human genes are all caps; Others are mouse genes

qPCR primer sequences			
Amplicon Name	Forward Primer Sequence (5' to 3')	Reverse Primer Sequence (5' to 3')	Amplicon size
ACTB	CCAACCGGAGAAGATGA	CCAGAGGCCATACAGGGATAG	97 bp
OCT4	CGTGAAGCTGGAGAAGGAGAAGCTG	AAGGGCCGAGCTTACACATGTC	246 bp
T	GGGTGGCTTCTGGAAAC	TTGAGAAITGTTCCGATGAG	172 bp
ETV2	GGTATGAAATGGGGCTGAGA	TCCAGCATGCTCTGCTGTC	230 bp
IRX5	GCTACAACCTGCACCTCCAG	ATGATCTTCGCCCCCTGGT	265 bp
XBP1	GGAGTTAACAGCAGCGCTTG	ACTGGGTCCAAGTTGCTCAG	248 bp
USF1	ATGGAGAGCACCAAGTCTGG	ACGACCTTAATCCGTGGTG	221 bp
MECOM	GCGAACACTATCCCCATGAA	GGCTTGCAGAACATCAGGAAA	261 bp
SOX2	AGAACCCAAGATGCACAAAC	GGGCAGCGTGTACTTATCCT	200 bp
NANOG	GAATGAAATCTAACAGGGCA	CTCTGGTAGGAAGAGTAAAGG	111 bp
Actb	CTAAGGCCAACCGTGAAAAG	ACCAGAGGCATACAGGACAA	104 bp
Pecam1	GAGCCAGCAGTATGAGGACCC	CTGTTGGCTTGGCTTTC	166 bp
Cdh5	AACGAACTGGATTCTCGGGG	TGAGGGTGAAGTTGCTGTC	231 bp
Oit3	CTCTCCACCACAGTCACCTT	CCACTCCCCATATGGT	150 bp
Tmem100	TGGCCTCTCTGGTAATGGAC	AGATTGGAAAGCTTGGTCA	63 bp
Foxf1	CCCTGTCGGCAGCATCT	TGTGAGTGAATCCGGGGATG	106 bp
Tspan7	ACCAAGTTATGGAGACTAACATGG	AGCAGCATGCAATCACT	83 bp
Hpgd	GCAGGGCTGAACATGAGA	AAACCAAGATAGGCTTCACTGA	83 bp
Acvr1l	CTGGGTGCTCTAGGCTTG	CCCAGGTCACTGTGAAAT	71 bp
Nkx2-3	AAGAGGAGGTTGAGCGAA	CAAAGACCTGAGCTTGGCG	156 bp
Col15a1	ACCTGCAAGGATCTCCACCA	ACTTGGCCCTTGGAGATTCAC	76 bp
Scgb3a1	TGAAGCTTACCAACACCTTC	CCATGAAGAAAGCAACACCA	72 bp
Pbx1	GCCAGACAGGAGGATAACGTG	CTGCCAACCTCATTAGCAC	83 bp
Meis2	AGACAAGGACGCAATCTATGG	GCTCGCAGCTCTAAAAACCC	68 bp
Igfbp5	CGCGAGACAAACAAAGATAG	TGTCTCATCTCAGGTGCA	237 bp
Slc1c1	GGAGGGCCATCTTACAGT	AGGGCACCAAGAACACCC	86 bp
Pglyrp1	CAATGTGAGCATTACCA	CCGTCTCTTCAATAAGGAA	75 bp
Slc2a1	TTACAGCGCTCCGTTCT	TCCCACAGCCAACATGAG	114 bp
Cpe	CTGGTCATCGAGCTGCTGA	GAATTGGGTGCTGTTGGATC	190 bp
Cyt1	ACACTACCTGTGCTGCT	TGGCCAGCACAGTAGTTA	214 b
Cfh	AAAAAACAAAGTGGCGAGA	GGAGGTGATGTCCTCATGGTC	75 bp
H19	CACTTTCCAAAGAGCTAACAC	GCTGGGTAGCACCATTCTT	95 bp
Spock2	agcccaagataaaggactg	agaaggctgttgtatcg	222 bp
Dcn	ACAAGTCTTGGGCTGAC	ACCCAGATCAGAACACTGCA	166 bp
Clec4g	TCTGAGCACCTACTGTCCA	TGCTCTGAGGCATTGTCC	88 bp
Dnase1l3	TCCCAGAGTGCAAAAGATGT	GTGTTGCTGCTGCTTGA	229 bp
Ehd3	TTTCGACAAACAAGCCATGG	CTCCACATCTCCATGCA	157 bp
Cd55	GGGACGATGAGTTCCAGT	CCCCTCGCATTATTCATT	78 bp

sgRNA sequences		
Name	Sequence (5' to 3')	Assay
IRX5 sgRNA1	CGTCATTTCGGGGCCCCGCA	KO construction
IRX5 sgRNA2	GGTATGTCGACATTAAACG	KO construction
MECOM sgRNA1	GGCGTGAGTGTACTAACCG	KO construction
MECOM sgRNA2	AACCTGATAACGTCATACG	KO construction
USF1 sgRNA1	TACATCACCTAGAAGTGTAG	KO construction
USF1 sgRNA2	TGATGACGACCTCTAATCCG	KO construction
XBP1 sgRNA1	CGGTGCGTAGTCTGGAGCTA	KO construction
XBP1 sgRNA2	GTCAATACGCCAGAACATCA	KO construction

Oligonucleotide sequences used for KO cell line construction			
Name	Upstream Sequence (5' to 3')	Downstream Sequence (5' to 3')	Assay
USF1_ssODN	TTCCAACACTTAATCTTCTCTCTTCACTCCCTGCCTTC	GGGAGGATTGGGAAGAGAAATAGCAGGATGCTGGGGACTAATT	
	CTAATCGCGTTACATAACTACGGTAAATGGCCCGCTGGCTG	AGAGGTCTGTCATCAAGAATGACAGCAACTATGGGGATICA	
		GGG	KO construction
MECOM_ssODN	AAGACTATCCCCATGAAACATGGCGCCGATATCCACGGATCGC	GGGAGGATTGGGAAGAGAAATAGCAGGATGCTGGGGACTAATG	
	GTTCATACTACCGTAAATGGCCCGCTGGCTG	ACGTATCAAGGTTGACCAGAGTGGACCAAGTCAA	KO construction
XBP1_ssODN	GCTGGGGCGTGGCGCTCGCGCGCGGGTGCCTGGCTG	GGGAGGATTGGGAAGAGAAATAGCAGGATGCTGGGGACTAATT	
	TCTGGAGATCGCGTTACATAACTACGGTAAATGGCCCGCTGGC	CTGGCGTATTGACTCTCAGATTAGGAGTAGGGATCATTCTGA	
		CT	KO construction
IRX5_ssODN	TCGTGCCCGCGTACAGCACCGCGTCATTGGGGCCCCATCGC	ggggaggatggggagagaatagcaggcatgtgggaCTAACGGGGCTGCG	
	GTTCATACTACCGTAAATGGCCCGCTGGCTG	TCGGTCCGGGACTTTCTAATTATTTAA	KO construction
CAGpro-Hyblintron	CTGATATCGCGTTACATAACTTACG	CTGAAAAAAAGTATTTCAGGC	selection cassette cloning
BSDR_CDS	GACCGGTACCATGCCAACGCTTGTCAA	GGATTCCCTACTGCAGCTGCAGGCCCT	selection cassette cloning
BGHpA	CCCGCAAGCCCCGGTGCTGACTGTGCTTCTAGTTGCC	CTATAG TCCCCAGCATGC	selection cassette cloning
USF1_KO_V1	CCCTATCCAAAGGAAAGCTTAT	GTACCTGGGTCTAGGGATTGAGT	Genotyping PCR
USF1_KO_V2	GTGCCAGCTTACTACCCATTIC	AGGACACCTCTGAACTTC	Genotyping PCR
XBP1_KO_V1	CACTCAGACTACGTGCACCTCT	ATGTCAAGCATCAAACAGATGG	Genotyping PCR
XBP1_KO_V2	AAGTACCTTGGCCAGGGATT	AGAACGAGAACCTTGGGGTCC	Genotyping PCR
MECOM_KO_V1	CATGAAGAGCAGAACACTATCCC	AAGCCAGGAAACTAACAAACA	Genotyping PCR
MECOM_KO_V2	CCTCATTCTACATCCACAGT	CTGCAGGTTATAAGGCATTC	Genotyping PCR
IRX5_KO_V1	GTCTTATCCGCAAGGGTACT	GTACTGGAGGTGCGAGTTGAG	Genotyping PCR
IRX5_KO_V2	GGGGTCTATTCAATGGATTAA	GATGGAAAAATAACTGCCAAGG	Genotyping PCR
KO_v_F3	TTCCCGGGCTAGAGCTCGCTGATCAGCCTC		Genotyping PCR
KO_v_R3	GGGCGTACTTGGCATATGAT		Genotyping PCR