Discovery of a primed endothelial progenitor that requires VEGF/ERK inhibition to complete vein differentiation

Lay Teng Ang[1,2,19,21,✉], Sherry Li Zheng[1,3,18,19], Kevin J. Liu[1,2,19], Anastasiia Masaltseva[1,3,4,19], June Winters[5,20], Sawan K. Jha[5,20], Qingqing Yin[1,3,20], Crystal Qian[1,3], Xiaochen Xiong[1,3,18], Amir Dailamy[1,3], Ellie Xi[1,2], Juan C. Alcocer[5], Daniel W. Sorensen[5,18], Richard She[6], Karina Smolyar[7], Dorota Szumska[8], Svanhild Nornes[8], Renata M. Martin[1,9,18], Benjamin J. Lesch[1,9,18], Nicole K. Restrepo[10], Wenfei Sun[11], Jonathan S. Weissman[7], Heiko Lickert[12,13,14], Matthew H. Porteus[1,9], Mark A. Skylar-Scott[4,15], Christian Mosimann[16], Saulius Sumanas[10], Sarah De Val[8,17], Kristy Red-Horse[1,5,✉], Kyle M. Loh[1,3,✉]

[1]Institute for Stem Cell Biology & Regenerative Medicine, Stanford University, Stanford, CA, USA
[2]Department of Urology, Stanford University, Stanford, CA, USA
[3]Department of Developmental Biology, Stanford University, Stanford, CA, USA
[4]Department of Bioengineering, Stanford University, Stanford, CA, USA
[5]Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA
[6]School of Biological Sciences, Nanyang Technological University, Singapore
[7]Whitehead Institute for Biomedical Research, Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA
[8]Institute of Developmental and Regenerative Medicine, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom
[9]Department of Pediatrics, Stanford University, Stanford, CA, USA
[10]Department of Pathology and Cell Biology, University of South Florida, Tampa, Florida, USA
[11]Division of Endocrinology, Department of Medicine, Stanford University, Stanford, CA, USA
[12]Institute of Diabetes and Regeneration Research, Helmholtz Center Munich, Neuherberg, Germany
[13]German Center for Diabetes Research, Neuherberg, Germany
[14]School of Medicine, Technical University of Munich, Munich, Germany
[15]Basic Science and Engineering Initiative, Children's Heart Center, Chan Zuckerberg Biohub, Stanford University, Stanford, CA, USA
[16]Section of Developmental Biology, Department of Pediatrics, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, CO, USA
[17]Ludwig Institute for Cancer Research Ltd, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[18]Current address: Stately Bio, Palo Alto, CA, USA (S.L.Z.), Department of Biology, Stanford University, Stanford, CA, USA (X.X.), Department of Biological Sciences, San José State University, San José, CA, USA (D.W.S.), Replay Bio, San Diego, CA, USA (R.M.M.), Department of Surgery, University of California San Francisco, San Francisco, USA (B.J.L.)
[19]These authors contributed equally to this work
[20]These authors contributed equally to this work
[21]Lead contact
✉Email: layteng@stanford.edu (L.T.A.), kredhors@stanford.edu (K.R.H.), kyleloh@stanford.edu (K.M.L.)

SUMMARY

Extracellular signals and cell-fate trajectories during vein development remain elusive, despite trailblazing insights into artery development. Here we exploit human pluripotent stem cell differentiation and mouse embryology to present a model that answers longstanding questions: vein endothelial cell (EC) differentiation unfolds in two steps driven by opposing extracellular signals. First, VEGF differentiates mesoderm into "primed" ECs, newly-defined progenitors that co-express certain arterial (SOX17) and venous (APLNR) markers. Second, primed ECs execute vein differentiation upon VEGF/ERK inhibition; however, upon VEGF activation they can instead form artery ECs. The arteriovenous plasticity of primed ECs was supported by intersectional lineage tracing. Future venous genes including *NR2F2* harbor poised chromatin in primed ECs, but are only transcribed upon VEGF/ERK inhibition. SOXF transcription factors, including SOX17, confer primed ECs with vein differentiation competence. Collectively, this two-step vein differentiation model—entailing primed EC intermediates and VEGF/ERK inhibition to trigger vein differentiation—has implications for VEGF-modulating therapies.

INTRODUCTION

Endothelial cells (ECs) comprise the inner lining of blood vessels, and there are multiple subtypes of ECs, including artery, vein, capillary, and lymphatic ECs; each executes specialized functions and expresses different genes[1-10]. For instance, artery ECs express NOTCH ligands that specify the smooth muscle cells that ensconce arteries[11]. By contrast, vein ECs can construct valves that prevent backward blood flow[12]. The prevailing model suggests that artery ECs express SOXF family transcription factors (SOX7, SOX17, and SOX18) that impart arterial identity[13-19]. Conversely, vein ECs express the transcription factor NR2F2 and transmembrane proteins APLNR/APJ and FLRT2[20-26].

Pathbreaking work has discovered key precepts of artery EC development[3-10]: VEGF/ERK activation[27-29], NOTCH activation[30-34], and cell-cycle arrest[23,35,36] all drive arterialization. Nevertheless, many mysteries continue to surround the development of vein ECs[6,8]. For instance, one might assume that arteries and veins should arise contemporaneously, as both are required to form a functional circulatory system. Nevertheless, during the development of the first intraembryonic blood vessels, arteries emerge earlier than veins, as shown by anatomy and molecular markers in mouse and zebrafish embryos[24,37-42]. Why veins should emerge after arteries remains a longstanding curiosity. This piques questions about the earliest steps of venous development and the exact identity of the progenitor cells that build veins.

The extracellular signals that specify venous identity also warrant further investigation[6,8]. The transcription factor NR2F2 is a master regulator of vein EC identity[22], but the upstream signals that ignite NR2F2 expression remain largely unknown[6,8]. VEGF is a master regulator of EC identity[4,43-47], but has paradoxical roles in venous development. On one hand, VEGF is *required* for vein EC formation, as ECs are absent in *Vegfr2*$^{-/-}$ mouse embryos[45]. On the other hand, VEGF also *represses* vein formation, as it instructs arterial identity in zebrafish embryos[27,28]. How can VEGF both promote and inhibit vein EC specification during vascular development?

Here we introduce a conceptual model of human vein EC differentiation that addresses certain mysteries surrounding vein development. We unveil two separable steps of vein differentiation, driven by opposing extracellular signals. First, VEGF *activation* differentiates lateral mesoderm

into a newly discovered "primed" EC intermediate, which co-expresses "arterial" marker SOX17 and "venous" marker APLNR. Primed ECs harbor the competence to generate both vein and artery ECs *in vitro*, consistent with broad arteriovenous plasticity in the vascular system[10]. Intersectional genetic tracing supports the notion that *Sox17*+ *Aplnr*+ ECs form both vein and artery ECs *in vivo*. Within primed ECs, future venous and arterial genes are poised at the chromatin level, but subsequently VEGF/ERK *inhibition* installs fully-fledged venous identity, and converts these poised venous genes into an active state.

The discovery of primed ECs may explain several longstanding curiosities in vascular development, including why veins tend to emerge after arteries *in vivo*[24,37-42]. Our model also reconciles why VEGF/ERK signaling can paradoxically both promote[45] and inhibit[27-29] vein EC development: the same signal executes both functions, at different times. The timing of an extracellular signal is of paramount importance, as it is interpreted in a temporally dynamic way. Our discovery that VEGF/ERK inhibition is required for vein EC specification also has potential implications for VEGF pathway inhibitors that are clinically used to block tumor angiogenesis[46,47].

RESULTS

Discovery of "primed" ECs that arise during human vascular differentiation *in vitro*

At 24 hour increments during hPSC differentiation into artery and vein ECs[48], we mapped gene expression (through single-cell [scRNAseq] and bulk-population RNA-sequencing), chromatin accessibility (using OmniATACseq[49]), and cell-surface marker expression (through high-throughput screening of 332 cell-surface markers) (**Fig. 1A**, **Fig. S1A-E**, **Table S1-S2**). We created an interactive web browser to explore scRNAseq data of each differentiation stage (https://anglab.shinyapps.io/artery-vein-scrna-seq/).

Investigating stepwise molecular changes during venous differentiation in detail, we discovered an unexpected intermediate with unique gene expression, chromatin, and cell-surface marker signatures—which we term "primed" ECs—that preceded the emergence of vein ECs. Day-2 lateral mesoderm expressed the angioblast transcription factor *ETV2*[50], but upon 24 hours of further differentiation, *ETV2* was silenced, whereupon EC surface markers (e.g., *CD144/VE-CADHERIN*) became expressed by day 3 (**Fig. 1B-D**, **Fig. S2A-B**). However, mature venous markers *NR2F2* and *FLRT2* were only expressed by day 4 of vein differentiation (**Fig. 1C-D, Fig. S2A-B**).

What is the identity of these incipient day-3 "primed" ECs emerging prior to expression of fully-fledged venous markers by day 4? Curiously, CD144+ primed ECs co-expressed the "arterial" marker *SOX17*[13-17] and "venous" marker *APLNR*[20,21] (**Fig. 1B,D**). By contrast, cells subjected to arterial differentiation solely expressed *SOX17*, but not *APLNR*, as expected[13-17,20-26] (**Fig. 1B,D**). We confirmed that single primed ECs co-expressed SOX17 and *APLNR*, as shown by combined *in situ* hybridization and immunostaining (**Fig. 1E**) and scRNAseq (**Fig. S2C**). However, primed ECs did not express additional markers of arterial (*GJA4/CX37*, *UNC5B*, *DLL4, MECOM, HEY1, EFNB2, CXCR4*, or *IGFBP3*), venous (*NR2F2*, *FLRT2*, or *NRP2*), or angioblast (*ETV2*) identity[23,24] (**Fig. 1D**, **Fig. S2B,D**). High-throughput profiling of cell-surface markers revealed that primed ECs did not express markers of artery ECs (CXCR4+ DLL4+) or vein ECs (CD73+ CD317+)[51-53], suggesting that primed ECs can be classified as neither arterial

nor venous (**Fig. S3A-F**, **Table S3**). Having discovered primed ECs, we subsequently explored the extracellular signals that drive entry into, and exit from, this intermediate state.

Two steps of vein EC differentiation *in vitro*: VEGF/ERK activation, followed by inhibition

What extracellular signals specify venous identity? We tested different combinations of the developmental signals shown in **Fig. 2A**. This revealed that VEGF *activation* was required to differentiate day-2 lateral mesoderm into day-3 primed ECs, alongside inhibition of the artery-specifying TGFβ and NOTCH pathways[30-34,48] (**Fig. 2B**). During this first 24-hour interval, VEGF/ERK activation was crucial to generate ECs; in its absence, EC formation *in vitro* was significantly impaired (**Fig. 2B**, **Fig. S4A**). The requirement for VEGF to differentiate lateral mesoderm into primed ECs is consistent with how VEGF is required for EC emergence *in vivo*[27,45].

While VEGF was essential to generate day-3 primed ECs, our combinatorial testing of developmental signals unexpectedly revealed that VEGF/ERK *inhibition* was necessary for their further progression into day-4 vein ECs (**Fig. 2C**, **Fig. S4B**). Upon VEGF/ERK inhibition, venous markers *NR2F2* and *FLRT2* sharply increased, and primed EC marker *SOX17* was silenced (**Fig. 2D-E**, **Fig. S2A**). 12 hours of VEGF/ERK inhibition sufficed to significantly activate *NR2F2* expression (**Fig. 2D**). By contrast, continued VEGF activation at this stage was arterializing, and repressed venous markers (**Fig. 2C-D**), as observed in zebrafish embryos[27,28]. The requirement for VEGF/ERK inhibition to confer venous identity was surprising, as VEGF is typically thought to be broadly required for all ECs[4,27,45]. Nevertheless, after cells have acquired endothelial identity, we found that VEGF was dispensable for cells to remain endothelial (**Fig. 2C**).

While primed ECs differentiated into vein ECs upon VEGF/ERK inhibition, they could instead adopt artery EC identity if challenged with artery-inducing signals, such as VEGF activation and PI3K inhibition[48,54] (**Fig. S4C**). The competence of primed ECs to form both vein and artery ECs *in vitro* reflects known arteriovenous plasticity within the vasculature *in vivo*[10].

Given that primed ECs harbor vein differentiation competence, we hypothesized that venous genes might be poised for future activation at the chromatin level. To test this hypothesis, we mapped the genome-wide distribution of H3K4me1, H3K4me3, and H3K27me3 using CUT&RUN[55]. In primed ECs, future venous genes such as *NR2F2*, *NT5E*, and *MAF* were not yet expressed, but nevertheless their promoter regions were bivalently marked by activation-associated H3K4me3 and repression-associated H3K27me3 (**Fig. 2F**, **Fig. S4D**). Such promoter bivalency can denote genes poised for future activation in anticipation of future developmental decisions[56]. VEGF/ERK inhibition transitioned the promoter elements of these venous genes from bivalency (H3K4me3+ H3K27me3+) to full activation (H3K4me3+ H3K27me3-), which was accompanied by transcription (**Fig. 2F**, **Fig. S4D**). Likewise, the promoter elements of future arterial genes were bivalently marked in primed ECs, foreshadowing arterial differentiation competence (**Fig. S4E**).

We thus propose two sequential steps of vein differentiation driven by temporally dynamic VEGF *activation* to specify primed ECs, followed by VEGF/ERK *inhibition* to induce vein ECs (**Fig. 2A**). This temporally-dynamic signaling switch reconciles why VEGF/ERK signaling has

been reported to both promote[45] and inhibit[27-29] vein formation *in vivo*: the same signal can do both, at different times.

Comparison of differentiation methods suggests that VEGF/ERK inhibition is integral to venous identity

Despite widespread success in differentiating hPSCs towards artery ECs, it has generally proven more challenging to generate vein ECs[6]. We hypothesized that difficulties in generating vein ECs might be attributable to the temporally-dynamic effects of VEGF: current differentiation methods typically entail prolonged VEGF activation[51-53,57-59], which might inadvertently deter venous differentiation. To this end, we compared 8 hPSC differentiation methods by collating published scRNAseq datasets and generating new ones[48,58-62] (**Fig. 3A**, **Fig. S5A-B**). What is the diversity and arteriovenous character of cells emerging from each protocol?

All 8 differentiation methods generated ECs and mesenchymal cells in varying proportions. The two differentiation protocols employed in this study[48] yielded the highest EC percentages (84.0-96.5%), with the remaining non-ECs comprising mesenchymal cells (**Fig. 3A**, **Fig. S5C**, **Fig. S6**). Certain differentiation protocols[60,61] yielded a wider spectrum of non-ECs, including blood, kidney, heart, and liver (**Fig. 3A**, **Fig. S5C**, **Fig. S6**). To minimize the emergence of non-ECs, we developed improved protocols to generate artery and vein ECs at higher purity and scale, yielding up to 654 million artery or vein ECs in a single batch (**Fig. S7A-F**).

Next, we quantified the arteriovenous character of ECs emerging from each *in vitro* differentiation protocol, using *in vivo*-defined arteriovenous marker signatures[24]. Multiple protocols generated artery ECs—perhaps reflecting the pervasive *in vitro* use of VEGF (**Fig. S5B**), which is arterializing *in vivo*[27,28]—although the arterial marker score differed between protocols (**Fig. 3B**, **Fig. S6**). The artery EC differentiation protocol employed in this study[48] imparted the strongest arterial identity (**Fig. 3B**; $P<0.0001$), potentially attributable to the explicit inhibition of vein-inducing signal PI3K[29,63] to sharpen arterial identity. Only two differentiation protocols[48,58] predominantly generated vein ECs, consistent with how it has proven more challenging to instate venous identity *in vitro*[6]. The protocol employed in this study[48]—which conferred the strongest venous identity (**Fig. 3C**; $P<0.0001$)—entailed VEGF-induced generation of primed ECs, followed by VEGF/ERK inhibition to specify vein ECs. Other methods relying on sustained VEGF activation[58] less effectively induced venous character (**Fig. 3C**, **Fig. S6**). After generating vein ECs, we developed enhanced methods to expand them for at least 6 days *in vitro* while preserving venous identity (**Fig. S7G-L**).

In summary, this comparison of differentiation protocols suggests that (1) transition through primed EC intermediates and (2) temporally dynamic VEGF/ERK modulation are critical to maximize vein EC generation *in vitro*. Prolonged VEGF activation is the cornerstone of extant EC differentiation protocols (**Fig. S5B**), but temporally dynamic VEGF/ERK activation, followed by inhibition, may prove decisive in imparting venous identity *in vitro*.

Sox17+ Aplnr+ primed ECs exist in early embryos and contribute to veins

Having discovered *SOX17*+ *APLNR*+ primed ECs *in vitro*, we investigated their existence and fate *in vivo*. Sox17 marks arteries *in vivo*[13-17], but others suggest that it is more broadly

expressed in the vasculature[64,65]. Whole mount immunostaining revealed that Sox17 marked most, if not all, Erg+ ECs in the E8.5 mouse embryo, including those in the dorsal aorta and the vitelline vein (**Fig. 4A-B**, **Fig. S8A**). In E9.5 mouse embryos, Sox17 was expressed by both the dorsal aorta and cardinal vein (**Fig. 4C**). Indeed, permanent labeling of *Sox17*+ cells and all of their progeny with *Sox17-Cre* resulted in both artery and vein ECs being lineage traced (**Fig. S8B-C**), consistent with past observations[66,67]. Despite the prominent role of Sox17 in artery development[13-17], these lineage tracing and immunostaining results demonstrate that early vein ECs also transiently express Sox17, although it is turned off later in the mature cardinal vein by E10.5[16,17]. Sox17 is thus expressed by most ECs in early vascular development, and these cells construct most of the nascent vasculature.

Sox17 and *Aplnr* were co-expressed in at least two vascular beds of the E9.5-E10 mouse embryo: the vascular plexus and early cardinal vein. First, the vascular plexus represents a plastic progenitor of both arteries and veins[10], and we found that the head and limb plexus expressed both Sox17 and *Aplnr* (**Fig. S8D**). Second, the common cardinal vein—the earliest intraembryonic vein[37]—also co-expressed Sox17 and *Aplnr*. This was independently confirmed by two different means to detect *Aplnr* expression: *in situ* hybridization (**Fig. 4D**) and Aplnr-CreER activity[20] (**Fig. S8D-E**). While the early cardinal vein anatomically appears as a vein, from a molecular perspective it transiently expresses Sox17 and *Aplnr*; it thus mirrors primed ECs *in vitro*, prior to downregulating "arterial" marker Sox17[16,17]. Suggestive of evolutionary conservation of this primed state, we found that *sox7* (a zebrafish *sox17* homolog[68]) was co-expressed alongside *aplnrb* in the zebrafish secondary vascular field (**Fig. 4E**), a plastic progenitor field capable of generating both arteries and veins[69].

Finally, we employed intersectional lineage tracing to test whether *Sox17*+ *Aplnr*+ ECs contribute to both artery and vein ECs. *Sox17-Cre*[66] and *Aplnr-DreER*[70] driver mice were crossed with an intersectional reporter mouse[71]. In this approach, *GFP* expression denotes both Cre- and Dre-mediated recombination[71], reflecting previous expression of *Sox17* and *Aplnr*. Tamoxifen injection at E8.5 led to GFP labeling of the E12.5 common cardinal vein, supporting the hypothesis that primed ECs form vein ECs *in vivo*; artery ECs were also labeled at lower frequency (**Fig. 4H**). Taken together, *Sox17*+ *Aplnr*+ ECs contribute to both veins and arteries in mouse embryos, and cells with molecular features of primed ECs may be evolutionarily conserved in zebrafish embryos.

SOXF transcription factors are required for both artery and vein differentiation

Our discovery that SOX17 marks primed ECs *in vivo* and *in vitro* was surprising, as the prevailing view is that the SOXF transcription factor (TF) family, including SOX7, SOX17, and SOX18, imparts arterial identity[13-19]. Primed ECs express all three *SOXF* TFs (**Fig. 5A**). Beyond artery identity, are SOXF TFs also required for human vein EC differentiation? To answer this question, we first engineered hPSCs to express an enhanced CRISPR interference (CRISPRi) effector (dCas9-ZIM3 KRAB)[72]. hPSCs were transduced with sgRNAs targeting *SOX7*, *SOX17,* and/or *SOX18*, and then subject to EC differentiation (**Fig. 5B**). Individual CRISPRi knockdown of either *SOX7*, *SOX17*, or *SOX18* modestly affected arteriovenous identity (**Fig. S9A-B**), consistent with genetic redundancy between SOXF TFs[7,68,73]. We thus simultaneously knocked down all three *SOXF* TFs using CRISPRi with high efficiency (**Fig. S9C-D**). As expected[13-19],

triple *SOXF* knockdown abrogated artery differentiation, with near-complete loss of arterial markers *CXCR4* and *DLL4* (**Fig. 5C-D**, **Fig. S9E-F**, **Table S5**).

Strikingly, triple *SOXF* knockdown also strongly impaired vein differentiation: venous markers *NR2F2*[22], *APLNR*[20,21] and *NT5E/CD73*[51-53] were reduced (**Fig. 5E**, **Fig. S9E**, **Table S6**). Beyond individual markers, triple *SOXF* knockdown significantly impaired the venous transcriptional signature more broadly, as computed from known venous marker panels[24] (**Fig. 5F-G**), and downregulated gene sets associated with blood vessel development, migration, and junction formation (**Fig. S9G**). Upon *SOXF* loss, cells retained pan-EC marker *PECAM1* (**Fig. 5C**, **Fig. S9C**), but selectively downregulated *VE-CADHERIN* (**Fig. S9E,H**), thus mirroring zebrafish results[74].

To investigate why SOXF TFs are required for venous differentiation, we mapped the genome-wide binding of SOX17 in primed ECs using CUT&RUN[55]. SOX17 directly bound the *APLNR* and *NR2F2* promoter elements in primed ECs, prior to venous differentiation and predating *NR2F2* transcription (**Fig. 5H**, **Fig. S9H**). Taken together, SOXF TFs directly bind to, and are required to activate, venous genes during human vein EC differentiation. This builds on the roles of SOXF TFs in mouse and zebrafish vein development[74-76]. We show that while SOXF is required for the specification of both artery and vein ECs, it regulates different genes in each: it is necessary for arterial gene expression in artery ECs, but is required for venous gene expression in vein ECs.

Accessible chromatin landscapes link VEGF/ERK inhibition to venous identity

We reasoned that TF motifs enriched in either artery- or vein-specific genomic regulatory elements might illuminate regulators of arteriovenous identity in an unbiased way. To this end, we mapped accessible chromatin in FACS-purified hPSC-derived artery and vein ECs using OmniATACseq[49] (**Fig. 6A**, **Table S7**). AP1 and ETS TFs are transcriptional effectors of ERK signaling[77], and remarkably their motifs comprised the top 15 TF motifs enriched in artery-specific regulatory elements, and were correspondingly depleted in vein-specific regulatory elements ($P<10^{-1080}$ and $P<10^{-958}$, respectively; **Fig. 6A**, **Table S8-S9**). At the mRNA level, AP1 (*JUNB*) and ETS (*ETV1*, *ETV4*, and *ETV5*) family TFs were upregulated in artery, relative to vein, ECs (**Fig. 6B**). As a positive control, motifs of known arterial (e.g., SOXF) and venous (e.g., NR2F, MAF) TF families were also reciprocally enriched in artery- or vein-specific regulatory elements (**Fig. 6A-B**, **Table S8-S9**). In summary, the accessible chromatin landscape of artery ECs is dominated by ERK transcriptional effectors, whereas vein EC accessible chromatin is depleted of such ERK effectors. The accessible chromatin landscapes of artery and vein ECs reveal how extracellular signals shape chromatin, and link VEGF/ERK inhibition to venous identity.

Chromatin profiling illuminated a wealth of regulatory elements with arterial- or venous-specific accessibility (**Fig. 6A**). In addition to known arterial enhancer elements in the *DLL4* and *CXCR4* loci[78,79] (**Fig. 6C-G**), our chromatin data revealed a novel, arterial-specific enhancer 13 kB upstream of *SEMA3G* gene, which encodes an arterially expressed guidance cue required for vascular pathfinding[80,81] (**Fig. 6H**). Transgenic reporter assays delineated that this novel *Sema3g* enhancer was active in arteries, but not veins, within mouse[82,83] and zebrafish embryos, attesting to its arterial specificity (**Fig. 6I, Fig. S10A**). Taken together, hPSC-derived

artery and vein ECs are distinguished by a wealth of arteriovenous regulatory elements at the chromatin level. Our genomic resource provides a means to identify additional such elements (**Table S7**, **Fig. S10B-C**), as exemplified by the arterial *SEMA3G* enhancer.

DISCUSSION

Curiously, despite incisive insights into mechanisms of artery EC development[3-10,23,27-36], longstanding mysteries have surrounded vein EC development[6,8]. What are the early embryonic progenitors that build veins? Why do veins often emerge after arteries *in vivo*[24,37-42]? What are the upstream extracellular signals that specify venous identity? Here we discover primed ECs that co-express the "arterial" marker *Sox17* and "venous" marker *Aplnr*, and can form vein ECs *in vivo* and *in vitro*. Consistent with widespread arteriovenous plasticity within the vasculature[10], primed ECs are not lineage-committed, as they also competent to generate artery ECs *in vivo* and *in vitro*. Collectively, the concept of primed ECs unifies several observations concerning endothelial development, including why the first intraembryonic veins generally emerge after arteries and why the earliest veins transiently "mis-express" certain arterial markers[24,37-42]. We further discover that VEGF/ERK inhibition is crucial for primed ECs to differentiate into vein ECs.

VEGF/ERK activation, followed by inhibition, drives vein EC differentiation

We present evidence for a two-step model for vein development *in vitro*: VEGF/ERK *activation* is initially required to endow endothelial identity during the conversion of lateral mesoderm into primed ECs, but 24 hours later, VEGF/ERK *inhibition* is subsequently required for vein EC specification. This reconciles the paradox of why VEGF/ERK both promotes[45] and represses[27-29] vein ECs *in vivo*: the same signal executes both roles, but at different times. The idea that VEGF/ERK inhibition is required for vein specification may be surprising, as VEGF activation is often construed as being broadly necessary for EC identity[4,27,45]. However, timing is paramount: while VEGF induces ECs, we suggest that once a cell has acquired EC identity, it now re-interprets VEGF/ERK inhibition to signify that it should adopt vein fate. Indeed, in other developmental venues, the same extracellular signal can specify different cellular identities at distinct times[84]. We propose that VEGF is a double-edged sword in venous differentiation: it induces ECs but subsequently represses venous identity. The temporally dynamic effects of VEGF likely contribute to past challenges in generating vein ECs *in vitro*. Indeed, VEGF is ubiquitous in EC differentiation and culture media, but our findings suggest that it may inadvertently suppress venous identity in certain contexts.

Our results may also pertain to the resistance of tumor vasculature against VEGF inhibitor therapies designed to block tumor angiogenesis[46,47]. If VEGF/ERK inhibition instructs ECs to adopt venous identity in certain contexts, this may represent a possible mechanism contributing to VEGF inhibitor resistance.

Why do major intraembryonic arteries precede veins *in vivo*? The answer may lie in how VEGF induces both endothelial and arterial identity[27-29,45]. Perhaps cells can contemporaneously acquire endothelial and arterial identity[37] under the command of VEGF, but vein development must be parsed into two distinct steps wherein endothelial identity is first instated, followed by VEGF/ERK inhibition to specify veins. At the chromatin level, future venous genes—such as *NR2F2*—are in a poised state in primed ECs, but are rapidly upregulated upon VEGF/ERK

inhibition. This fills a vital missing link in our knowledge of the upstream extracellular signals that ignite *NR2F2* expression[6,8].

In other developmental contexts, the rich detail afforded by scRNAseq suggests that differentiation paths are continua[85,86], piquing the philosophical question of where one cell state (e.g., primed) ends and the other (e.g., vein) begins. Nevertheless, primed ECs represent an intermediate step in arteriovenous differentiation, as they differ from vein ECs in multiple ways. First, they are functionally induced by opposing signals: VEGF/ERK activation vs. inhibition, respectively. Second, primed ECs do not express later-stage venous markers such as *NR2F2*, but harbor bivalent chromatin at future venous genes, implying developmental poising. Third, primed ECs lack venous cell-surface markers CD73 and CD317.

SOXF transcription factors are required for vein and artery EC specification

While the early cardinal vein anatomically appears as a vein, we find that it molecularly co-expresses Sox17 and *Aplnr*. The early cardinal vein may thus transiently resemble primed ECs, prior to downregulating Sox17 at later stages[16,17]. Indeed, most if not all ECs temporarily express Sox17 in the mouse embryo, as shown by immunostaining and *Sox17-Cre* lineage tracing. Consistent with how the early vasculature broadly expresses Sox17 *in vivo*, simultaneous CRISPRi knockdown of all *SOXF* family members (*SOX7*, *SOX17*, and *SOX18*) compromised both human artery and vein EC differentiation *in vitro*.

Our findings support longstanding observations that SOXF transcription factors are required for arterial specification[13-19], but additionally suggest important roles in venous identity. Further evidence supports our conclusions. First, both arteries and veins are morphologically lost in the *Sox7*-/- *Sox17*-/- *Sox18*-/- mouse retinal vasculature[76]. Second, double *sox7* and *sox18* knockdown abrogates vein development in zebrafish[74,75]. Third, SOXF motifs reside in not just arterial enhancers, but also in pan-endothelial enhancers, implying that SOXF has broader roles in endothelial identity beyond arterialization[79]. Taken together, SOXF transcription factors are expressed in primed ECs, and are crucial for human vein EC differentiation.

While SOXF is necessary for the specification of artery and vein ECs, it activates different genes in each. Our CRISPRi studies showed that SOXF upregulates arterial genes in artery ECs, whereas it drives expression of certain venous genes (e.g., *NR2F2* and *APLNR*) in vein ECs. Indeed, TFs execute cell-type-specific roles[87]. Could differing SOXF levels contribute to these outcomes? Artery ECs express higher levels of SOX17 than primed ECs (**Fig. 1D**). Perhaps higher vs. lower SOXF levels respectively bias ECs to arterial vs. venous identities, but wholesale SOXF depletion—as we achieved by CRISPRi—compromises both. SOXF TFs may also interact with differing cofactors in either artery vs. vein ECs to direct distinct transcriptional programs.

Taken together, we propose a two-step model for vein differentiation, wherein VEGF activation initially differentiates mesoderm into primed ECs, and subsequently, VEGF/ERK inhibition specifies vein ECs. This provides a new perspective on vein development and has practical ramifications, as hPSC differentiation into primed ECs represents a critical step to generate vein ECs.

LIMITATIONS OF THE STUDY

First, VEGF/ERK is crucial for incipient EC specification, but how does it subsequently become repressed to permit vein EC specification *in vivo*? When and where VEGF ligands are expressed is the most parsimonious explanation for this temporally dynamic VEGF/ERK signaling switch. Initially, VEGF differentiates lateral mesoderm into ECs *in vivo*[27,45,88]. Subsequently, future cardinal vein ECs become separated from the VEGF source by several hundred microns, as shown in *Xenopus*[89], likely explaining why VEGF/ERK signaling becomes attenuated to drive vein specification.

Second, what are the molecular mechanisms through which VEGF/ERK inhibition endows venous identity? In other cell-types, ERK2 transcriptionally pauses RNA polymerase II on future developmental genes[90]. It will be interesting to test if RNA polymerase II elongation is unleashed upon VEGF/ERK inhibition in primed ECs, driving *NR2F2* upregulation.

Third, our intersectional lineage tracing revealed that *Sox17*+ *Aplnr*+ ECs contribute to vein ECs, but not every vein EC was labeled. This likely reflecting incomplete recombination efficiency inherent to intersectional lineage tracing, or alternative sources of vein ECs. An open question remains whether *all* vein ECs emerge via primed EC intermediates, or whether alternative differentiation paths can also culminate in venous identity. This might be addressable in the future through unbiased CRISPR barcoding[91,92].

ACKNOWLEDGEMENTS

We thank Valerie Park, Liying Ou, Aaron McCarty, Catherine Carswell-Crumpton, the Stanford Institute for Stem Cell Biology & Regenerative Medicine, and the Stanford Stem Cell FACS Core Facility for infrastructure support. We thank Sophie Astrof for sharing *Aplnr-DreER* mice, and Ryan Corces, Arjun Rajan, Zhainib Amir-Ugokwe, Siva Vijayakumar, Len Pennacchio, Axel Visel, Stacey Lee, Avishek Ganguly, Yan Yang, Yimiao Qu, Alicia Wong, and Julie Chen for assistance. Genomics analyses were supported by the Stanford Diabetes Genomics Analysis Core Facility (NIH P30DK116074). This work was supported by NIH Director's Early Independence Award DP5OD024558 (K.M.L.), NIH R01HL128503 (K.R.-H.), NIH R01HL153005 (S.S.), NIH R01HL168097 (C.M.), Breakthrough T1D Northern California Center of Excellence (K.M.L.), Additional Ventures Expansion Award (K.M.L., L.T.A.), Siebel Stem Cell Institute (L.T.A.), Ludwig Cancer Research (K.M.L., S.D.V.), Stanford Maternal and Child Health Research Institute Chambers Family Foundation Innovation Grant (K.M.L., L.T.A., K.R.-H.), Advanced Research Projects Agency for Health (ARPA-H) under award number AY1AX000002 (M.A.S.-S., K.M.L.), Helmholtz Society (H.L.), German Research Foundation (H.L.), European Research Council (H.L.), British Heart Foundation FS/SBSRF/22/31037 and RE/18/3/34214 (S.D.V.), Amaranth Foundation (K.M.L.), and Anonymous and Gilbert families (K.M.L.). S.L.Z. was supported by NSF and Stanford Graduate Fellowships. A.M. was supported by the Stanford Graduate Fellowship. R.M.M. was supported by NIH T32GM119995. C.M. is the The Helen and Arthur E. Johnson Chair For The Cardiac Research Director. L.T.A. is an Additional Ventures Catalyst to Independence Fellow and a Bladder Cancer Advocacy Network Career Development Awardee. K.R.-H. is a Howard Hughes Medical Institute Investigator. K.M.L. is a Human Frontier Science Program Young Investigator (RGY0069/2019), Packard Foundation Fellow, Pew Scholar, Baxter Foundation Faculty Scholar, and The Anthony DiGenova Endowed Faculty Scholar. Content is solely the

responsibility of the authors and does not necessarily represent the official views of the NIH, ARPA-H, or other funding organizations.

AUTHOR CONTRIBUTIONS

L.T.A., S.L.Z., K.J.L., A.M., Q.Y., C.Q., X.X., and K.M.L. differentiated and characterized hPSCs. S.L.Z., J.W., S.K.J., J.A., D.W.S., and K.R.-H. analyzed mouse embryos. L.T.A., S.L.Z., K.J.L., A.M., X.X., A.D., E.X., W.S., C.M., S.D.V., and K.M.L. performed genomics analyses. R.S., K.S., R.M.M., B.J.L., J.S.W., H.L., M.A.S.-S., and M.H.P. developed and provided reagents and advice. N.K.R. and S.S. stained zebrafish embryos. D.S., S.N., and S.D.V. analyzed enhancer transgenic embryos. L.T.A., K.R.-H., and K.M.L. supervised the study, with input from S.L.Z., K.J.L., and A.M.

DECLARATION OF INTERESTS

Stanford University has filed patent applications related to endothelial differentiation. S.L.Z. and R.M.M. are presently employed in the biotechnology industry, but contributed to this work while affiliated with Stanford University.

REFERENCES

1. Rafii, S., Butler, J.M., and Ding, B.-S. (2016). Angiocrine functions of organ-specific endothelial cells. Nature *529*, 316-325. 10.1038/nature17040.
2. Augustin, H.G., and Koh, G.Y. (2017). Organotypic vasculature: From descriptive heterogeneity to functional pathophysiology. Science *357*. 10.1126/science.aal2379.
3. Potente, M., and Mäkinen, T. (2017). Vascular heterogeneity and specialization in development and disease. Nature Reviews Molecular Cell Biology *18*, 477-494. 10.1038/nrm.2017.36.
4. Coultas, L., Chawengsaksophak, K., and Rossant, J. (2005). Endothelial cells and VEGF in vascular development. Nature *438*, 937-945. 10.1038/nature04479.
5. De Val, S., and Black, B.L. (2009). Transcriptional control of endothelial cell development. Dev Cell *16*, 180-195. 10.1016/j.devcel.2009.01.014.
6. Loh, K.M., and Ang, L.T. (2023). Building human artery and vein endothelial cells from pluripotent stem cells, and enduring mysteries surrounding arteriovenous development. Semin Cell Dev Biol. 10.1016/j.semcdb.2023.06.004.
7. Fish, J.E., and Wythe, J.D. (2015). The molecular regulation of arteriovenous specification and maintenance. Developmental Dynamics *244*, 391-409. 10.1002/dvdy.24252.
8. Lin, F.-J., Tsai, M.-J., and Tsai, S.Y. (2007). Artery and vein formation: a tug of war between different forces. EMBO Rep *8*, 920-924.
9. Fang, J., and Hirschi, K. (2019). Molecular regulation of arteriovenous endothelial cell specification. F1000Research *8*, 1208. 10.12688/f1000research.16701.1.
10. Trimm, E., and Red-Horse, K. (2023). Vascular endothelial cell development and diversity. Nat Rev Cardiol *20*, 197-210. 10.1038/s41569-022-00770-1.
11. High, F.A., Lu, M.M., Pear, W.S., Loomes, K.M., Kaestner, K.H., and Epstein, J.A. (2008). Endothelial expression of the Notch ligand Jagged1 is required for vascular smooth muscle development. Proc Natl Acad Sci U S A *105*, 1955-1959. 10.1073/pnas.0709663105.

12. Bazigou, E., and Makinen, T. (2013). Flow control in our vessels: vascular valves make sure there is no way back. Cell Mol Life Sci *70*, 1055-1066. 10.1007/s00018-012-1110-6.
13. Corada, M., Orsenigo, F., Morini, M.F., Pitulescu, M.E., Bhat, G., Nyqvist, D., Breviario, F., Conti, V., Briot, A., Iruela-Arispe, M.L., et al. (2013). Sox17 is indispensable for acquisition and maintenance of arterial identity. Nature Communications *4*, 2609. 10.1038/ncomms3609.
14. Sakamoto, Y., Hara, K., Kanai-Azuma, M., Matsui, T., Miura, Y., Tsunekawa, N., Kurohmaru, M., Saijoh, Y., Koopman, P., and Kanai, Y. (2007). Redundant roles of Sox17 and Sox18 in early cardiovascular development of mouse embryos. Biochem Biophys Res Commun *360*, 539-544.
15. Matsui, T., Kanai-Azuma, M., Hara, K., Matoba, S., Hiramatsu, R., Kawakami, H., Kurohmaru, M., Koopman, P., and Kanai, Y. (2006). Redundant roles of Sox17 and Sox18 in postnatal angiogenesis in mice. J Cell Sci *119*, 3513-3526.
16. Hosking, B., Francois, M., Wilhelm, D., Orsenigo, F., Caprini, A., Svingen, T., Tutt, D., Davidson, T., Browne, C., Dejana, E., and Koopman, P. (2009). Sox7 and Sox17 are strain-specific modifiers of the lymphangiogenic defects caused by Sox18 dysfunction in mice. Development *136*, 2385-2391. 10.1242/dev.034827.
17. Clarke, R.L., Yzaguirre, A.D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W.S., Speck, N.A., and Keller, G. (2013). The expression of Sox17 identifies and regulates haemogenic endothelium. Nature Cell Biology *15*, 502-510.
18. Herpers, R., van de Kamp, E., Duckers, H.J., and Schulte-Merker, S. (2008). Redundant roles for sox7 and sox18 in arteriovenous specification in zebrafish. Circ Res *102*, 12-15. 10.1161/CIRCRESAHA.107.166066.
19. Pendeville, H., Winandy, M., Manfroid, I., Nivelles, O., Motte, P., Pasque, V., Peers, B., Struman, I., Martial, J.A., and Voz, M.L. (2008). Zebrafish Sox7 and Sox18 function together to control arterial-venous identity. Dev Biol *317*, 405-416. 10.1016/j.ydbio.2008.01.028.
20. Chen, H.I., Sharma, B., Akerberg, B.N., Numi, H.J., Kivela, R., Saharinen, P., Aghajanian, H., McKay, A.S., Bogard, P.E., Chang, A.H., et al. (2014). The sinus venosus contributes to coronary vasculature through VEGFC-stimulated angiogenesis. Development *141*, 4500-4512. 10.1242/dev.113639.
21. Saint-Geniez, M., Argence, C.B., Knibiehler, B., and Audigier, Y. (2003). The msr/apj gene encoding the apelin receptor is an early and specific marker of the venous phenotype in the retinal vasculature. Gene Expr Patterns *3*, 467-472. 10.1016/s1567-133x(03)00062-0.
22. You, L.-R., Lin, F.-J., Lee, C.T., DeMayo, F.J., Tsai, M.-J., and Tsai, S.Y. (2005). Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. Nature *435*, 98-104.
23. Su, T., Stanley, G., Sinha, R., Amato, G.D.x., Das, S., Rhee, S., Chang, A.H., Poduri, A., Raftrey, B., Dinh, T.T., et al. (2018). Single-cell analysis of early progenitor cells that build coronary arteries. Nature, 1-29. 10.1038/s41586-018-0288-7.
24. Hou, S., Li, Z., Dong, J., Gao, Y., Chang, Z., Ding, X., Li, S., Li, Y., Zeng, Y., Xin, Q., et al. (2022). Heterogeneity in endothelial cells and widespread venous arterialization during early vascular development in mammals. Cell Res *32*, 333-348. 10.1038/s41422-022-00615-z.

25. Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., Bian, L., et al. (2019). Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. Cell Research, 1-14. 10.1038/s41422-019-0228-6.
26. Kalucka, J., de Rooij, L.P.M.H., Goveia, J., Rohlenova, K., Dumas, S.J., Meta, E., Conchinha, N.V., Taverna, F., Teuwen, L.-A., Veys, K., et al. (2020). Single-Cell Transcriptome Atlas of Murine Endothelial Cells. Cell, 1-37. 10.1016/j.cell.2020.01.015.
27. Casie Chetty, S., Rost, M.S., Enriquez, J.R., Schumacher, J.A., Baltrunaite, K., Rossi, A., Stainier, D.Y.R., and Sumanas, S. (2017). Vegf signaling promotes vascular endothelial differentiation by modulating etv2 expression. Developmental Biology *424*, 147-161. 10.1016/j.ydbio.2017.03.005.
28. Lawson, N.D., Vogel, A.M., and Weinstein, B.M. (2002). sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. Developmental Cell *3*, 127-136.
29. Hong, C.C., Peterson, Q.P., Hong, J.-Y., and Peterson, R.T. (2006). Artery/vein specification is governed by opposing phosphatidylinositol-3 kinase and MAP kinase/ERK signaling. Curr Biol *16*, 1366-1372.
30. Lawson, N.D., Scheer, N., Pham, V.N., Kim, C.H., Chitnis, A.B., Campos-Ortega, J.A., and Weinstein, B.M. (2001). Notch signaling is required for arterial-venous differentiation during embryonic vascular development. Development *128*, 3675-3683.
31. Krebs, L.T., Shutter, J.R., Tanigaki, K., Honjo, T., Stark, K.L., and Gridley, T. (2004). Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants. Genes & Development *18*, 2469-2473. 10.1101/gad.1239204.
32. Duarte, A., Hirashima, M., Benedito, R., Trindade, A., Diniz, P., Bekman, E., Costa, L., Henrique, D., and Rossant, J. (2004). Dosage-sensitive requirement for mouse Dll4 in artery development. Genes & Development *18*, 2474-2478. 10.1101/gad.1239004.
33. Gale, N.W., Dominguez, M.G., Noguera, I., Pan, L., Hughes, V., Valenzuela, D.M., Murphy, A.J., Adams, N.C., Lin, H.C., Holash, J., et al. (2004). Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. Proceedings of the National Academy of Sciences of the United States of America *101*, 15949-15954.
34. Fischer, A., Schumacher, N., Maier, M., Sendtner, M., and Gessler, M. (2004). The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. Genes Dev *18*, 901-911. 10.1101/gad.291004.
35. Luo, W., Garcia-Gonzalez, I., n, M.F.x.E.n.-C.x.F., Casquero-Garcia, V.x.F.n., oz, M.S.S.-M.x.F., hleder, S.M.x.F., Garcia-Ortega, L., Andrade, J., Potente, M., and Benedito, R. (2020). Arterialization requires the timely suppression of cell growth. Nature, 1-29. 10.1038/s41586-020-3018-x.
36. Fang, J.S., Coon, B.G., Gillis, N., Chen, Z., Qiu, J., Chittenden, T.W., Burt, J.M., Schwartz, M.A., and Hirschi, K.K. (2017). Shear-induced Notch-Cx37-p27 axis arrests endothelial cell cycle to enable arterial specification. Nat Commun *8*, 2149. 10.1038/s41467-017-01742-7.
37. Chong, D.C., Koo, Y., Xu, K., Fu, S., and Cleaver, O. (2011). Stepwise arteriovenous fate acquisition during mammalian vasculogenesis. Developmental Dynamics *240*, 2153-2165. 10.1002/dvdy.22706.
38. Drake, C.J., and Fleming, P.A. (2000). Vasculogenesis in the day 6.5 to 9.5 mouse embryo. Blood *95*, 1671-1679.

39. Kim, Y.H., Hu, H., Guevara-Gallardo, S., Lam, M.T., Fong, S.Y., and Wang, R.A. (2008). Artery and vein size is balanced by Notch and ephrin B2/EphB4 during angiogenesis. Development *135*, 3755-3764. 10.1242/dev.022475.
40. Herbert, S.P., Huisken, J., Kim, T.N., Feldman, M.E., Houseman, B.T., Wang, R.A., Shokat, K.M., and Stainier, D.Y.R. (2009). Arterial-venous segregation by selective cell sprouting: an alternative mode of blood vessel formation. Science (New York, NY) *326*, 294-298.
41. Lindskog, H., Kim, Y.H., Jelin, E.B., Kong, Y., Guevara-Gallardo, S., Kim, T.N., and Wang, R.A. (2014). Molecular identification of venous progenitors in the dorsal aorta reveals an aortic origin for the cardinal vein in mammals. Development *141*, 1120-1128. 10.1242/dev.101808.
42. Coffin, J.D., Harrison, J., Schwartz, S., and Heimark, R. (1991). Angioblast differentiation and morphogenesis of the vascular endothelium in the mouse embryo. Dev Biol *148*, 51-62. 10.1016/0012-1606(91)90316-u.
43. Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., et al. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. Nature *380*, 435-439. 10.1038/380435a0.
44. Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K.S., Powell-Braxton, L., Hillan, K.J., and Moore, M.W. (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. Nature *380*, 439-442. 10.1038/380439a0.
45. Shalaby, F., Rossant, J., Yamaguchi, T.P., Gertsenstein, M., Wu, X.F., Breitman, M.L., and Schuh, A.C. (1995). Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. Nature *376*, 62-66.
46. Patel, S.A., Nilsson, M.B., Le, X., Cascone, T., Jain, R.K., and Heymach, J.V. (2023). Molecular Mechanisms and Future Implications of VEGF/VEGFR in Cancer Therapy. Clin Cancer Res *29*, 30-39. 10.1158/1078-0432.CCR-22-1366.
47. Mori, T., Kumar, R.N.N., and Ferrara, N. (2024). Elucidating VEGF Biology: A Journey of Discovery and Clinical Translation. Arterioscler Thromb Vasc Biol *44*, 2361-2365. 10.1161/ATVBAHA.124.319574.
48. Ang, L.T., Nguyen, A.T., Liu, K.J., Chen, A., Xiong, X., Curtis, M., Martin, R.M., Raftry, B.C., Ng, C.Y., Vogel, U., et al. (2022). Generating human artery and vein cells from pluripotent stem cells highlights the arterial tropism of Nipah and Hendra viruses. Cell *185*, 2523-2541 e2530. 10.1016/j.cell.2022.05.024.
49. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nature Methods *528*, 142. 10.1038/nmeth.4396.
50. Koyano-Nakagawa, N., and Garry, D.J. (2017). Etv2 as an essential regulator of mesodermal lineage development. Cardiovascular Research *113*, 1294-1306. 10.1093/cvr/cvx133.
51. Ditadi, A., Sturgeon, C.M., Tober, J., Awong, G., Kennedy, M., Yzaguirre, A.D., Azzola, L., Ng, E.S., Stanley, E.G., French, D.L., et al. (2015). Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. Nature Cell Biology *17*, 580-591. 10.1038/ncb3161.

52. Park, M.A., Kumar, A., Jung, H.S., Uenishi, G., Moskvin, O.V., Thomson, J.A., and Slukvin, I.I. (2018). Activation of the Arterial Program Drives Development of Definitive Hemogenic Endothelium with Lymphoid Potential. Cell Reports *23*, 2467-2481. 10.1016/j.celrep.2018.04.092.
53. Gage, B.K., Liu, J.C., Innes, B.T., MacParland, S.A., McGilvray, I.D., Bader, G.D., and Keller, G.M. (2020). Generation of Functional Liver Sinusoidal Endothelial Cells from Human Pluripotent Stem-Cell-Derived Venous Angioblasts. Cell Stem Cell. 10.1016/j.stem.2020.06.007.
54. Amir Ugokwe, Z., Pyke, A.L., Trimm, E., Chakraborty, M., Fan, X., Ang, L.T., Loh, K.M., and Red-Horse, K. (2025). VEGF/ERK activation and PI3K inhibition together drive a vein-to-artery transition in an in vitro model of human angiogenesis. bioRxiv. 10.64898/2025.12.17.694993.
55. Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*. 10.7554/eLife.21856.
56. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.
57. Rosa, S., Praça, C., Pitrez, P.R., Gouveia, P.J., Aranguren, X.L., Ricotti, L., and Ferreira, L.S. (2019). Functional characterization of iPSC-derived arterial- and venous-like endothelial cells. Scientific Reports *9*, 3826-3815. 10.1038/s41598-019-40417-9.
58. Pan, Z., Yao, Q., Kong, W., Ma, X., Tian, L., Zhao, Y., Zhu, S., Chen, S., Sun, M., Liu, J., et al. (2024). Generation of iPSC-derived human venous endothelial cells for the modeling of vascular malformations and drug discovery. Cell Stem Cell. 10.1016/j.stem.2024.10.015.
59. Nikolova, M.T., He, Z., Seimiya, M., Jonsson, G., Cao, W., Okuda, R., Wimmer, R.A., Okamoto, R., Penninger, J.M., Camp, J.G., and Treutlein, B. (2025). Fate and state transitions during human blood vessel organoid development. Cell. 10.1016/j.cell.2025.03.037.
60. Paik, D.T., Tian, L., Lee, J., Sayed, N., Chen, I.Y., Rhee, S., Rhee, J.-W., Kim, Y., Wirka, R.C., Buikema, J.W., et al. (2018). Large-Scale Single-Cell RNA-Seq Reveals Molecular Signatures of Heterogeneous Populations of Human Induced Pluripotent Stem Cell-Derived Endothelial Cells. Circulation Research. 10.1161/CIRCRESAHA.118.312913.
61. Zhang, J., Chu, L.-F., Hou, Z., Schwartz, M.P., Hacker, T., Vickerman, V., Swanson, S., Leng, N., Nguyen, B.K., Elwell, A., et al. (2017). Functional characterization of human pluripotent stem cell-derived arterial endothelial cells. Proceedings of the National Academy of Sciences of the United States of America, 201702295. 10.1073/pnas.1702295114.
62. McCracken, I.R., Taylor, R.S., Kok, F.O., de la Cuesta, F., Dobie, R., Henderson, B.E.P., Mountford, J.C., Caudrillier, A., Henderson, N.C., Ponting, C.P., and Baker, A.H. (2019). Transcriptional dynamics of pluripotent stem cell-derived endothelial cell differentiation revealed by single-cell RNA sequencing. European Heart Journal *385*, 9963117. 10.1093/eurheartj/ehz351.
63. Sabata, H., Roca-Coll, A., Dengra, J.A., Gouveia, L., Martinez-Larrinaga, A., Collado-Remacha, A., Castillo, S.D., Castillo, E., Tisch, N., De Andrés-Laguillo, M., et al. (2025). Context-dependent response of endothelial cells to PIK3CA mutation. bioRxiv. 10.1101/2025.02.25.640041.

643 64. González-Hernández, S., Gomez, M.J., Sanchez-Cabo, F., Mendez-Ferrer, S., Munoz-
644 Canoves, P., and Isern, J. (2020). Sox17 Controls Emergence and Remodeling of
645 Nestin-Expressing Coronary Vessels. Circ Res *127*, e252-e270.
646 10.1161/CIRCRESAHA.120.317121.
647 65. Sharma, B., Ho, L., Ford, G.H., Chen, H.I., Goldstone, A.B., Woo, Y.J., Quertermous,
648 T., Reversade, B., and Red-Horse, K. (2017). Alternative Progenitor Cells Compensate
649 to Rebuild the Coronary Vasculature in Elabela- and Apj-Deficient Hearts. Dev Cell *42*,
650 655-666 e653. 10.1016/j.devcel.2017.08.008.
651 66. Engert, S., Liao, W.P., Burtscher, I., and Lickert, H. (2009). Sox17-2A-iCre: a knock-in
652 mouse line expressing Cre recombinase in endoderm and vascular endothelial cells.
653 Genesis *47*, 603-610.
654 67. Choi, E., Kraus, M.R.C., Lemaire, L.A., Yoshimoto, M., Vemula, S., Potter, L.A.,
655 Manduchi, E., Stoeckert, C.J., Grapin-Botton, A., and Magnuson, M.A. (2012). Dual
656 lineage-specific expression of Sox17 during mouse embryogenesis. Stem cells (Dayton,
657 Ohio) *30*, 2297-2308.
658 68. Lilly, A.J., Lacaud, G., and Kouskoff, V. (2017). SOXF transcription factors in
659 cardiovascular development. Semin Cell Dev Biol *63*, 50-57.
660 10.1016/j.semcdb.2016.07.021.
661 69. Metikala, S., Warkala, M., Casie Chetty, S., Chestnut, B., Rufin Florat, D., Plender, E.,
662 Nester, O., Koenig, A.L., Astrof, S., and Sumanas, S. (2022). Integration of vascular
663 progenitors into functional blood vessels represents a distinct mechanism of vascular
664 growth. Dev Cell *57*, 767-782 e766. 10.1016/j.devcel.2022.02.015.
665 70. Ramirez, A., Vyzas, C.A., Zhao, H., Eng, K., Degenhardt, K., and Astrof, S. (2024).
666 Buffering Mechanism in Aortic Arch Artery Formation and Congenital Heart Disease.
667 Circ Res *134*, e112-e132. 10.1161/CIRCRESAHA.123.322767.
668 71. Plummer, N.W., Evsyukova, I.Y., Robertson, S.D., de Marchena, J., Tucker, C.J., and
669 Jensen, P. (2015). Expanding the power of recombinase-based labeling to uncover
670 cellular diversity. Development *142*, 4385-4393. 10.1242/dev.129981.
671 72. Alerasool, N., Segal, D., Lee, H., and Taipale, M. (2020). An efficient KRAB domain for
672 CRISPRi applications in human cells. Nat Methods *17*, 1093-1096. 10.1038/s41592-
673 020-0966-x.
674 73. Payne, S., Neal, A., and De Val, S. (2024). Transcription factors regulating
675 vasculogenesis and angiogenesis. Dev Dyn *253*, 28-58. 10.1002/dvdy.575.
676 74. Cermenati, S., Moleri, S., Cimbro, S., Corti, P., Del Giacco, L., Amodeo, R., Dejana, E.,
677 Koopman, P., Cotelli, F., and Beltrame, M. (2008). Sox18 and Sox7 play redundant roles
678 in vascular development. Blood *111*, 2657-2666. 10.1182/blood-2007-07-100412.
679 75. Swift, M.R., Pham, V.N., Castranova, D., Bell, K., Poole, R.J., and Weinstein, B.M.
680 (2014). SoxF factors and Notch regulate nr2f2 gene expression during venous
681 differentiation in zebrafish. Dev Biol *390*, 116-125. 10.1016/j.ydbio.2014.03.018.
682 76. Zhou, Y., Williams, J., Smallwood, P.M., and Nathans, J. (2015). Sox7, Sox17, and
683 Sox18 Cooperatively Regulate Vascular Development in the Mouse Retina. PLoS One
684 *10*, e0143650. 10.1371/journal.pone.0143650.
685 77. Lavoie, H., Gagnon, J., and Therrien, M. (2020). ERK signalling: a master regulator of
686 cell behaviour, life and fate. Nat Rev Mol Cell Biol *21*, 607-632. 10.1038/s41580-020-
687 0255-7.

78. Sacilotto, N., Monteiro, R., Fritzsche, M., Becker, P.W., Sanchez-Del-Campo, L., Liu, K., Pinheiro, P., Ratnayaka, I., Davies, B., Goding, C.R., et al. (2013). Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development. Proceedings of the National Academy of Sciences of the United States of America *110*, 11893-11898. 10.1073/pnas.1300805110.
79. Nornes, S., Bruche, S., Adak, N., McCracken, I.R., and De Val, S. (2025). Evaluating the transcriptional regulators of arterial gene expression via a catalogue of characterized arterial enhancers. Elife *14*. 10.7554/eLife.102440.
80. Liu, X., Uemura, A., Fukushima, Y., Yoshida, Y., and Hirashima, M. (2016). Semaphorin 3G Provides a Repulsive Guidance Cue to Lymphatic Endothelial Cells via Neuropilin-2/PlexinD1. Cell Rep *17*, 2299-2311. 10.1016/j.celrep.2016.11.008.
81. Kutschera, S., Weber, H., Weick, A., De Smet, F., Genove, G., Takemoto, M., Prahst, C., Riedel, M., Mikelis, C., Baulande, S., et al. (2011). Differential endothelial transcriptomics identifies semaphorin 3G as a vascular class 3 semaphorin. Arterioscler Thromb Vasc Biol *31*, 151-159. 10.1161/ATVBAHA.110.215871.
82. Kosicki, M., Baltoumas, F.A., Kelman, G., Boverhof, J., Ong, Y., Cook, L.E., Dickel, D.E., Pavlopoulos, G.A., Pennacchio, L.A., and Visel, A. (2025). VISTA Enhancer browser: an updated database of tissue-specific developmental enhancers. Nucleic Acids Res *53*, D324-D330. 10.1093/nar/gkae940.
83. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88-92. 10.1093/nar/gkl822.
84. Fowler, J.L., Ang, L.T., and Loh, K.M. (2019). A critical look: Challenges in differentiating human pluripotent stem cells into desired cell types and organoids. Wiley Interdisciplinary Reviews: Developmental Biology *113*, 891-823. 10.1002/wdev.368.
85. Wagner, D.E., and Klein, A.M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. Nature Reviews Genetics, 1-18. 10.1038/s41576-020-0223-2.
86. Domcke, S., and Shendure, J. (2023). A reference cell tree will serve science better than a reference cell atlas. Cell *186*, 1103-1114. 10.1016/j.cell.2023.02.016.
87. Kim, S., and Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. Mol Cell *83*, 373-392. 10.1016/j.molcel.2022.12.032.
88. Rasmussen, T.L., Shi, X., Wallis, A., Kweon, J., Zirbes, K.M., Koyano-Nakagawa, N., and Garry, D.J. (2012). VEGF/Flk1 signaling cascade transactivates Etv2 gene expression. PLoS One *7*, e50103. 10.1371/journal.pone.0050103.
89. Cleaver, O., and Krieg, P.A. (1998). VEGF mediates angioblast migration during development of the dorsal aorta in Xenopus. Development *125*, 3905-3914. 10.1242/dev.125.19.3905.
90. Tee, W.W., Shen, S.S., Oksuz, O., Narendra, V., and Reinberg, D. (2014). Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. Cell *156*, 678-690. 10.1016/j.cell.2014.01.009.
91. Chan, M.M., Smith, Z.D., Grosswendt, S., Kretzmer, H., Norman, T.M., Adamson, B., Jost, M., Quinn, J.J., Yang, D., Jones, M.G., et al. (2019). Molecular recording of mammalian embryogenesis. Nature, 1. 10.1038/s41586-019-1184-5.
92. Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.C., Fujiwara, Y., Li, B.E., et al. (2020). An Engineered CRISPR-

Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell *181*, 1410-1422 e1427. 10.1016/j.cell.2020.04.048.

93. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e1821. 10.1016/j.cell.2019.05.031.

94. Calvanese, V., Capellera-Garcia, S., Ma, F., Fares, I., Liebscher, S., Ng, E.S., Ekstrand, S., Aguade-Gorgorio, J., Vavilina, A., Lefaudeux, D., et al. (2022). Mapping human haematopoietic stem cells from haemogenic endothelium to birth. Nature *604*, 534-540. 10.1038/s41586-022-04571-x.

95. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nature Neuroscience *13*, 133-140. 10.1038/nn.2467.

96. Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593-605. 10.1002/dvg.20335.

97. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550. 10.1073/pnas.0506580102.

98. Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

99. Kreitzer, F.R., Salomonis, N., Sheehan, A., Huang, M., Park, J.S., Spindler, M.J., Lizarraga, P., Weiss, W.A., So, P.L., and Conklin, B.R. (2013). A robust method to derive functional neural crest cells from human pluripotent stem cells. Am J Stem Cells *2*, 119-131.

100. Guna, A., Page, K.R., Replogle, J.M., Esantsi, T.K., Wang, M.L., Weissman, J.S., and Voorhees, R.M. (2023). A dual sgRNA library design to probe genetic modifiers using genome-wide CRISPRi screens. BMC Genomics *24*, 651. 10.1186/s12864-023-09754-y.

101. Goldstein, A.S., Drake, J.M., Burnes, D.L., Finley, D.S., Zhang, H., Reiter, R.E., Huang, J., and Witte, O.N. (2011). Purification and direct transformation of epithelial progenitor cells from primary human prostate. Nature Protocols *6*, 656-667. 10.1038/nprot.2011.317.

102. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682. 10.1038/nmeth.2019.

103. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529. 10.1016/j.cell.2021.04.048.

104. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications *8*, 14049. 10.1038/ncomms14049.

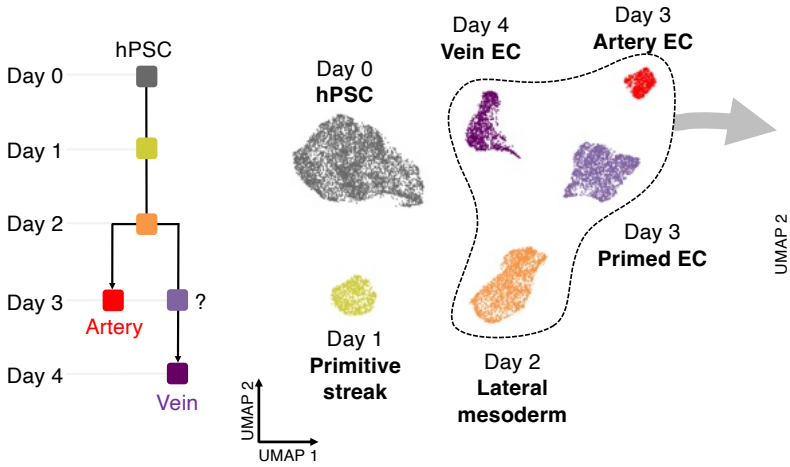
105. Rainer, J. (2025). EnsDb.Hsapiens.v86. https://bioconductor.org/packages/release/data/annotation/html/EnsDb.Hsapiens.v86.html.
106. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550. 10.1186/s13059-014-0550-8.
107. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287. 10.1089/omi.2011.0118.
108. Ouyang, J.F., Kamaraj, U.S., Cao, E.Y., and Rackham, O.J.L. (2021). ShinyCell: Simple and sharable visualisation of single-cell gene expression data. Bioinformatics. 10.1093/bioinformatics/btab209.
109. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
110. Krueger, F. (2012). Trim Galore. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
111. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology *34*, 525-527. 10.1038/nbt.3519.
112. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140. 10.1093/bioinformatics/btp616.
113. Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol *15*, R29. 10.1186/gb-2014-15-2-r29.
114. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47. 10.1093/nar/gkv007.
115. Wickham, H., François, R., Henry, L., and Müller, K. (2022). dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org.
116. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).
117. Wickham, H., Averick, M., Bryan, J., and Chang, W. (2019). Welcome to the Tidyverse. The Journal of Open Source Software. 10.21105/joss.01686.
118. Wingett, S.W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res *7*, 1338. 10.12688/f1000research.15931.2.
119. Gaspar, J.M. (2018). NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinformatics *19*, 536. 10.1186/s12859-018-2579-2.
120. Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics *30*, 2503-2505. 10.1093/bioinformatics/btu314.
121. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079. 10.1093/bioinformatics/btp352.

122. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165. 10.1093/nar/gkw257.
123. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res *4*, 1521. 10.12688/f1000research.7563.2.
124. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389-393. 10.1038/nature10730.
125. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.
126. Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., et al. (2011). Chemically defined conditions for human iPSC derivation and culture. Nature Methods *8*, 424-429.
127. Loh, K.M., Zheng, S.L., Liu, K.J., Yin, Q., Amir-Ugokwe, Z.A., Jha, S.K., Qi, Y., Wazny, V.K., Nguyen, A.T., Chen, A., et al. (2024). Protocol for efficient generation of human artery and vein endothelial cells from pluripotent stem cells. STAR Protoc *6*, 103494. 10.1016/j.xpro.2024.103494.
128. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26. 10.1038/nbt.1754.
129. Fowler, J.L., Zheng, S.L., Nguyen, A., Chen, A., Xiong, X., Chai, T., Chen, J.Y., Karigane, D., Banuelos, A.M., Niizuma, K., et al. (2024). Lineage-tracing hematopoietic stem cell origins in vivo to efficiently make human HLF+ HOXA+ hematopoietic progenitors from pluripotent stem cells. Dev Cell. 10.1016/j.devcel.2024.03.003.
130. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753-165122. 10.1242/dev.165753.
131. Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C Y.Y., Nichane, M., et al. (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell *14*, 237-252.
132. Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451-467. 10.1016/j.cell.2016.06.011.
133. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnology *37*, 38-44. 10.1038/nbt.4314.
134. Wimmer, R.A., Leopoldi, A., Aichinger, M., Wick, N., Hantusch, B., Novatchkova, M., Taubenschmid, J., mmerle, M.H.x.E., Esk, C., Bagley, J.A., et al. (2019). Human blood

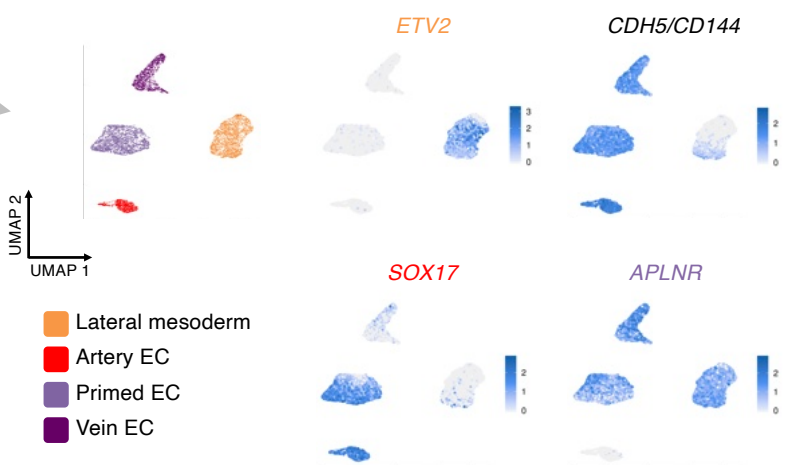
868 vessel organoids as a model of diabetic vasculopathy. Nature, 1-34. 10.1038/s41586-
869 018-0858-8.
870 135. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2.
871 Nature Methods *9*, 357-359. 10.1038/nmeth.1923.
872 136. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E.,
873 Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based
874 analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.
875 137. Gerstein, M.B., Kundaje, A.B., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu,
876 X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human
877 regulatory network derived from ENCODE data. Nature *488*, 91-100.
878 10.1038/nature11245.
879 138. Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields,
880 A.P., Park, C.Y., Corn, J.E., Kampmann, M., and Weissman, J.S. (2016). Compact and
881 highly active next-generation libraries for CRISPR-mediated gene repression and
882 activation. eLife *5*, 914. 10.7554/eLife.19760.
883 139. Martin, R.M., Ikeda, K., Cromer, M.K., Uchida, N., Nishimura, T., Romano, R., Tong,
884 A.J., Lemgart, V.T., Camarena, J., Pavel-Dinu, M., et al. (2019). Highly Efficient and
885 Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP
886 and AAV6 Donor-Mediated Homologous Recombination. Cell Stem Cell *24*, 821-
887 828.e825. 10.1016/j.stem.2019.04.001.
888 140. Chen, J., Zhang, X., DeLaughter, D.M., Trembley, M.A., Saifee, S., Xiao, F., Chen, J.,
889 Zhou, P., Seidman, C.E., Seidman, J.G., and Pu, W.T. (2024). Molecular and Spatial
890 Signatures of Mouse Embryonic Endothelial Cells at Single-Cell Resolution. Circ Res
891 *134*, 529-546. 10.1161/CIRCRESAHA.123.323956.

Figure 1

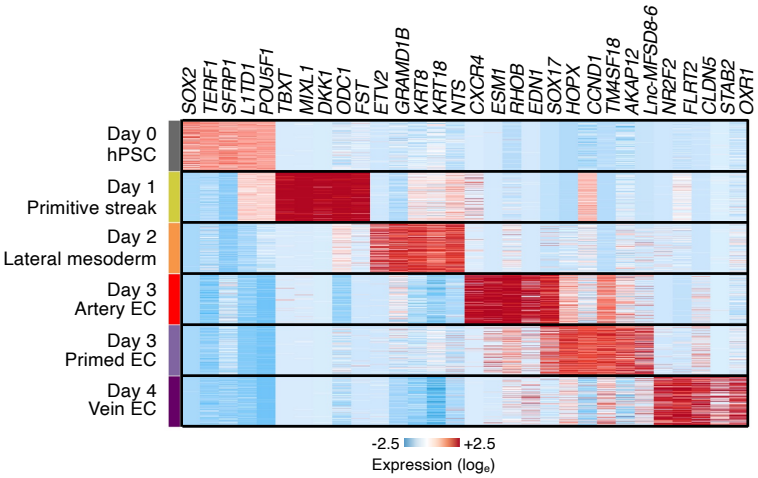
A scRNA-seq profiling of hPSC differentiation



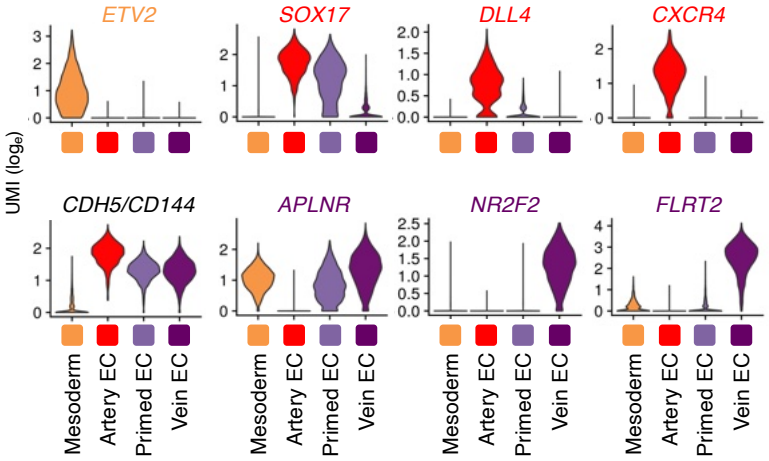
B Discovery of "primed" endothelial cells *in vitro*



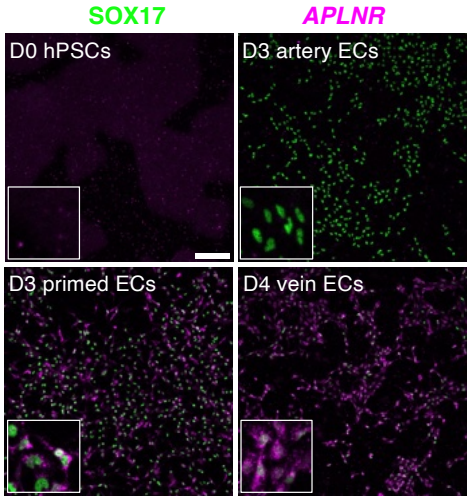
C Differentially expressed genes



D Expression of arterial and venous markers by primed ECs *in vitro*



E Primed ECs co-express SOX17 and *APLNR in vitro*



F Primed ECs express neither arterial nor venous surface markers

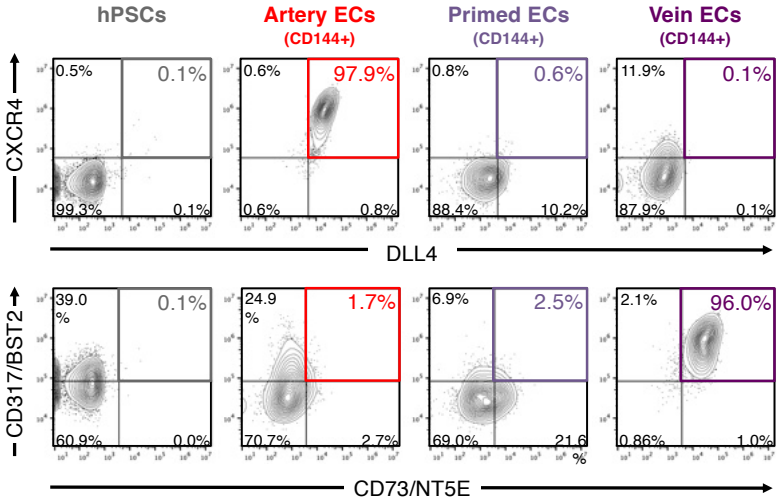
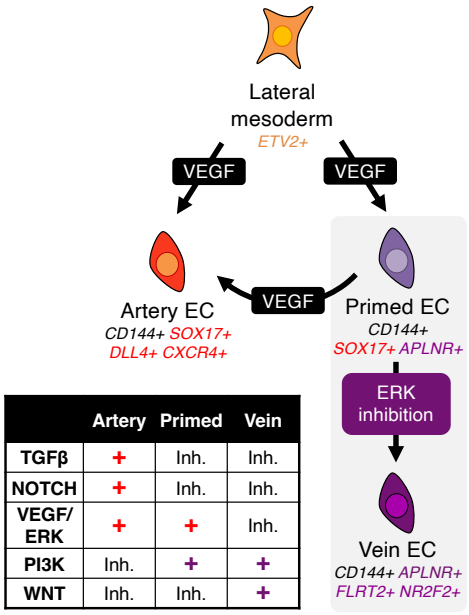


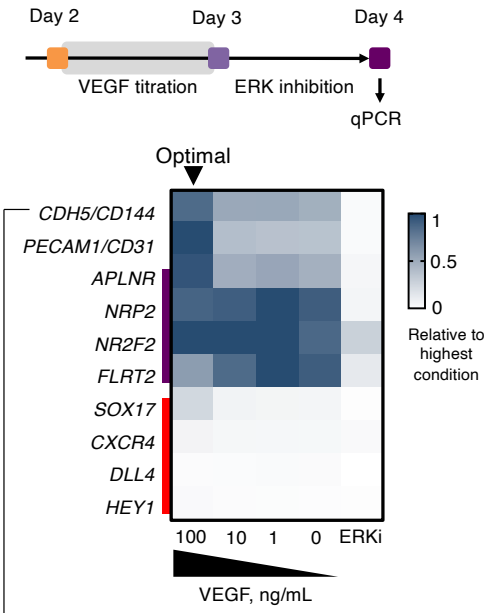
Figure 2

A Two vein differentiation steps: VEGF/ERK activation, then inhibition

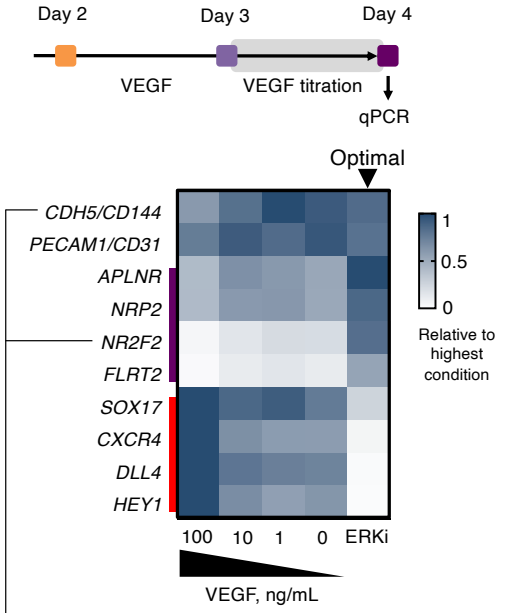


	Artery	Primed	Vein
TGFβ	+	Inh.	Inh.
NOTCH	+	Inh.	Inh.
VEGF/ ERK	+	+	Inh.
PI3K	Inh.	+	+
WNT	Inh.	Inh.	+

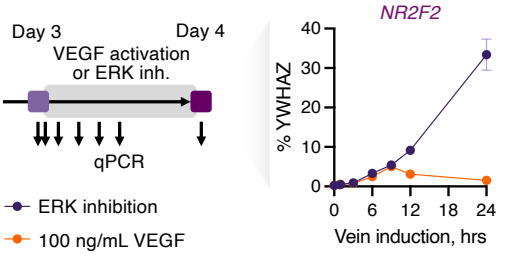
B VEGF/ERK activation on D2-3 interval is required for EC identity

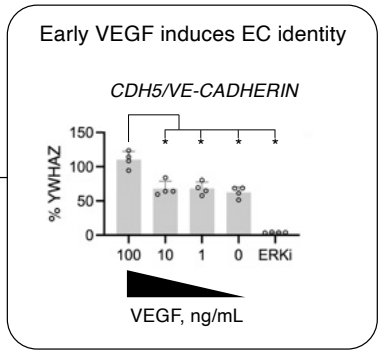


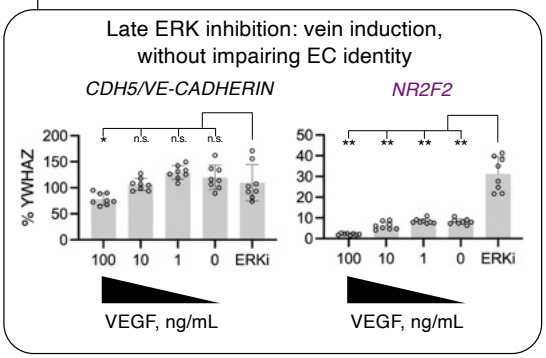
C VEGF/ERK inhibition on D3-4 interval is required for vein EC induction



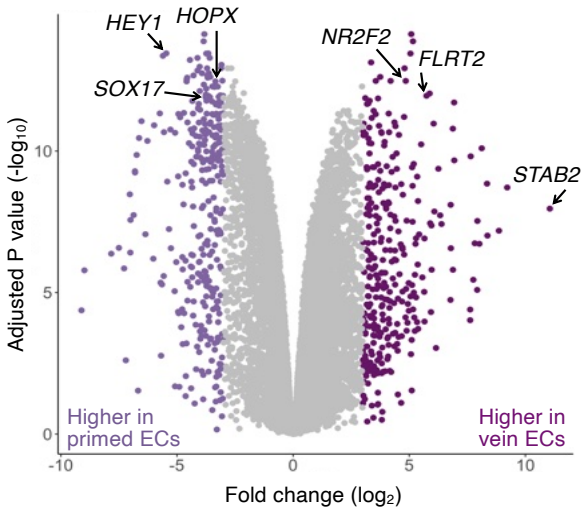
D ERK inhibition acutely induces vein markers







E Vein genes activated during primed to vein transition



F *NR2F2* promoter is bivalently marked in primed ECs

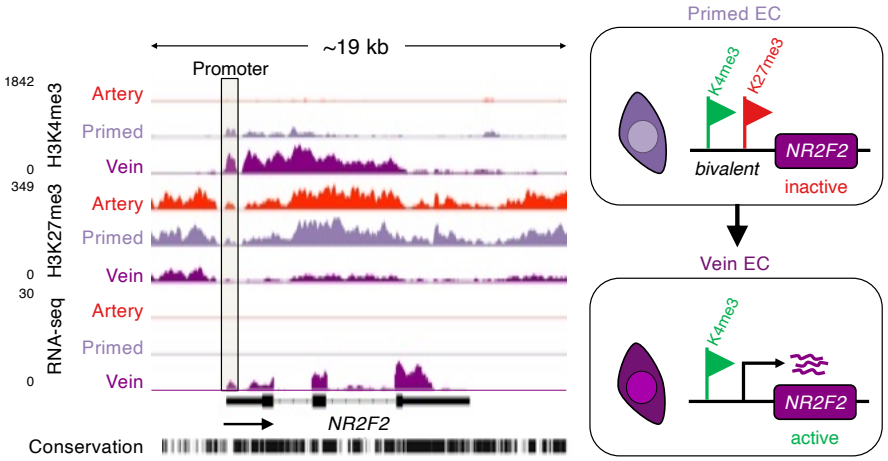
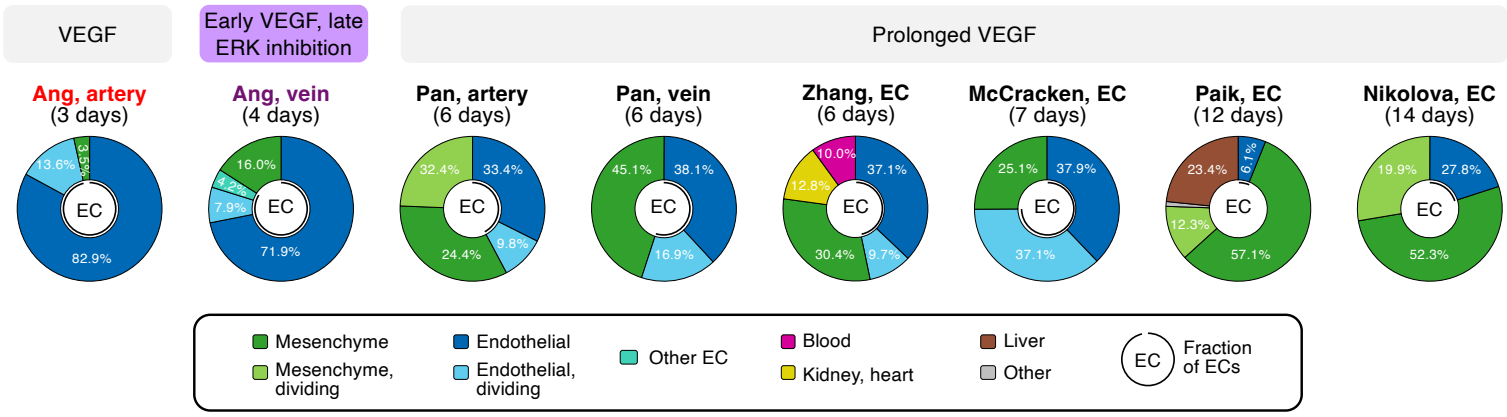
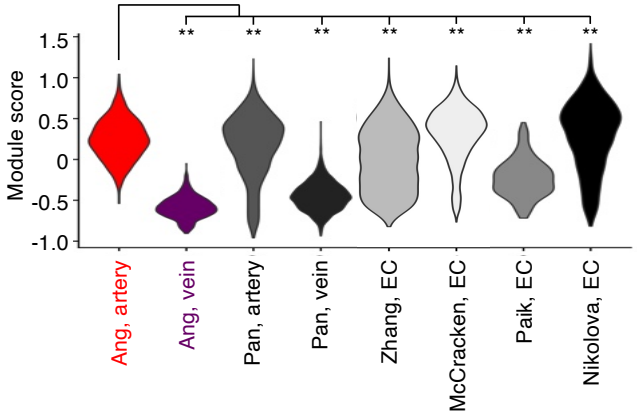


Figure 3

A Comparison of EC differentiation methods (scRNA-seq)



B Artery signature (scRNA-seq)



C Vein signature (scRNA-seq)

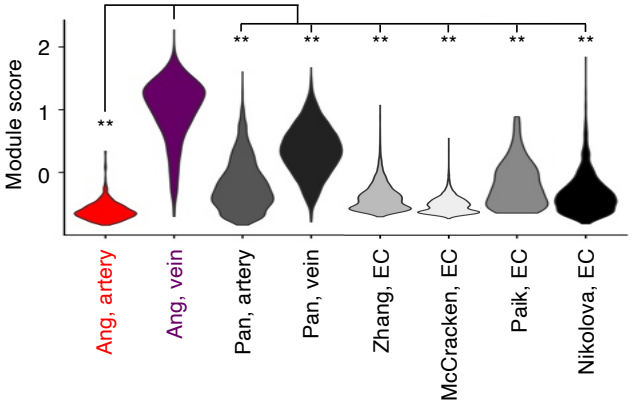
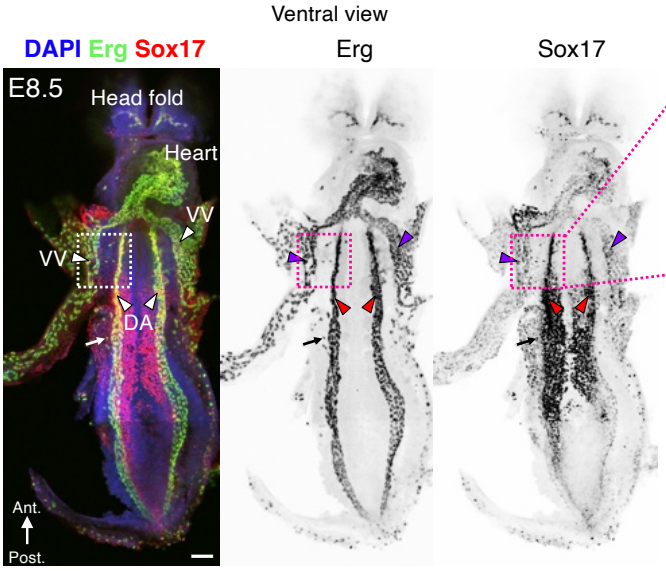
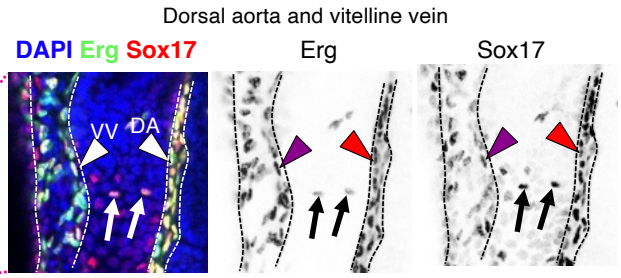


Figure 4

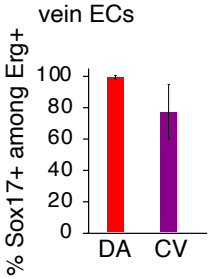
A Broad Sox17 expression in E8.5 mouse embryo ECs



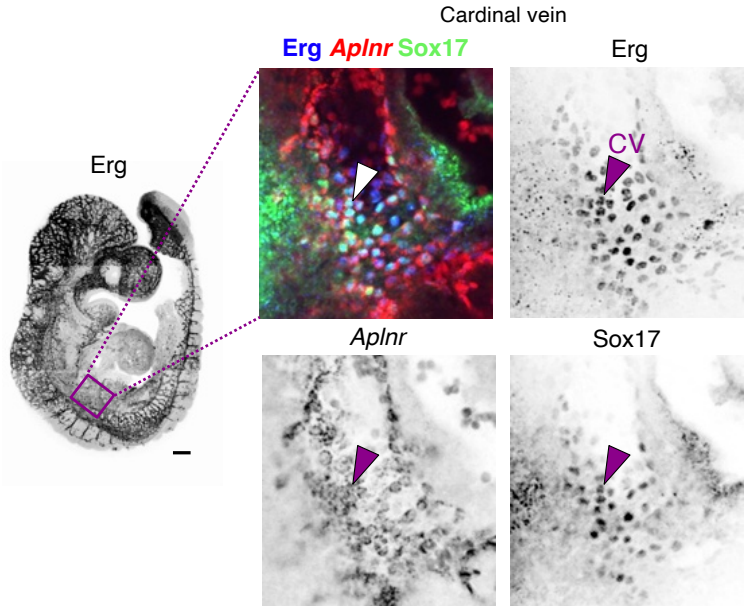
B Early artery and vein ECs at E8.5 express Sox17



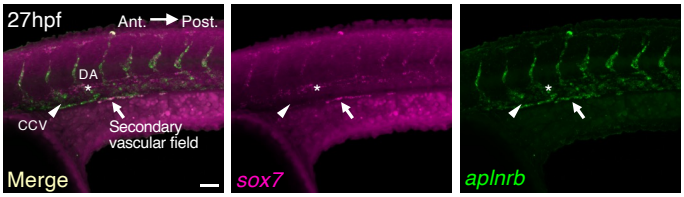
C Sox17 marks E9.5 artery and vein ECs



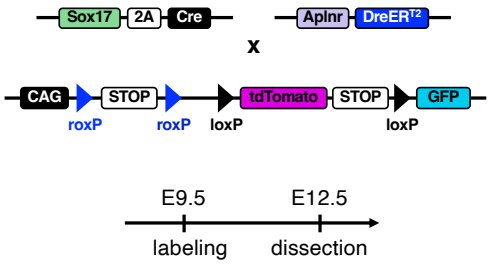
D E9.5 cardinal vein ECs co-express Sox17 and *Aplnr*

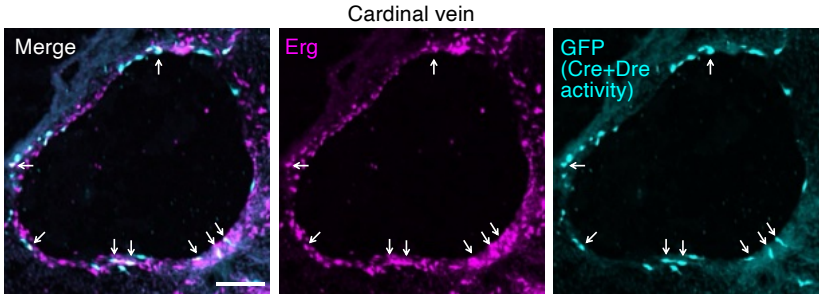


E *In situ* hybridization (27 hpf zebrafish embryo)



F *Sox17-Cre;Aplnr-DreERT2* intersectional lineage tracing (E12.5 mouse embryo)





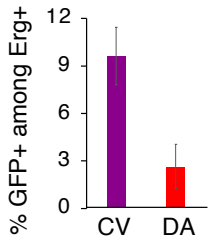
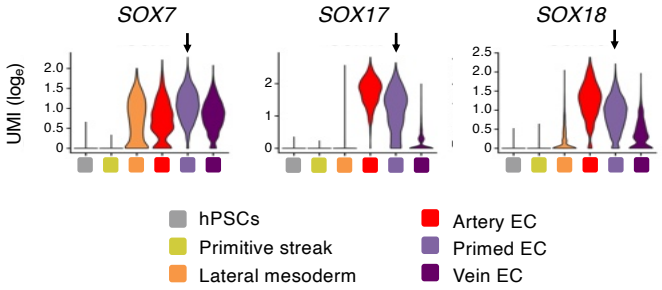
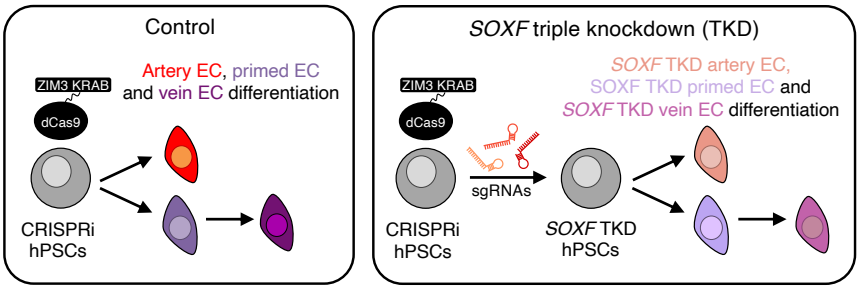


Figure 5

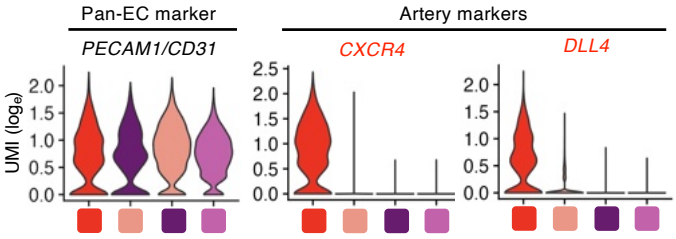
A Primed ECs express all *SOXF* family members



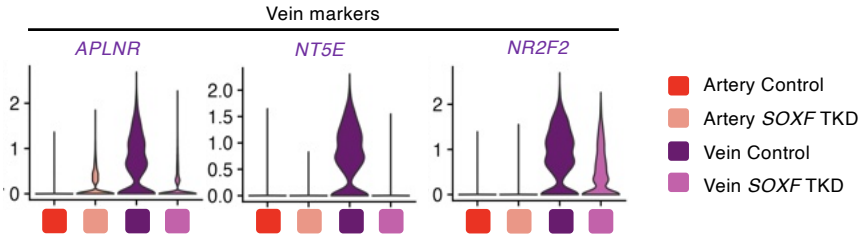
B Experimental strategy



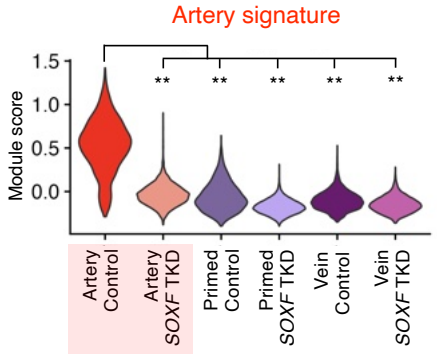
C *SOXF* genes required for artery identity



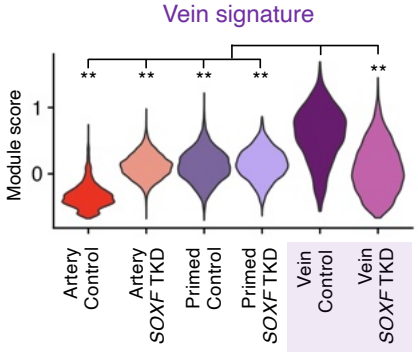
E *SOXF* genes required for vein identity



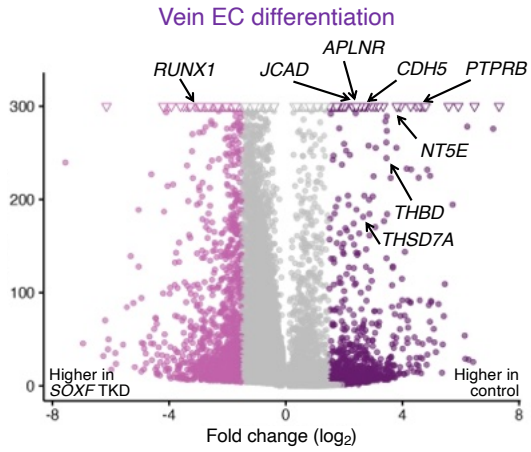
D *SOXF* genes required for artery signature



F *SOXF* genes required for vein signature



G *SOXF* loss abrogates vein EC differentiation



H SOX17 binds *APLNR* and *NR2F2* in primed ECs

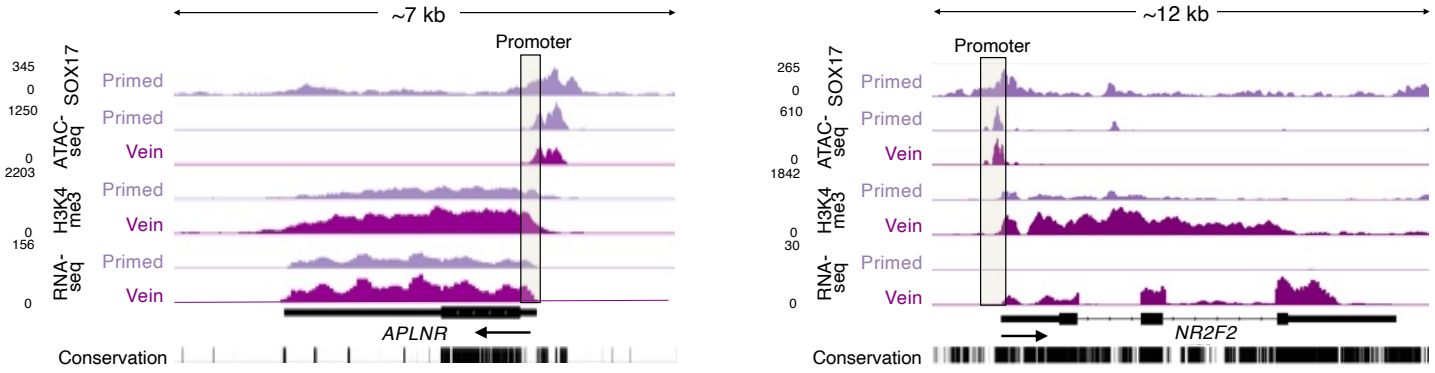
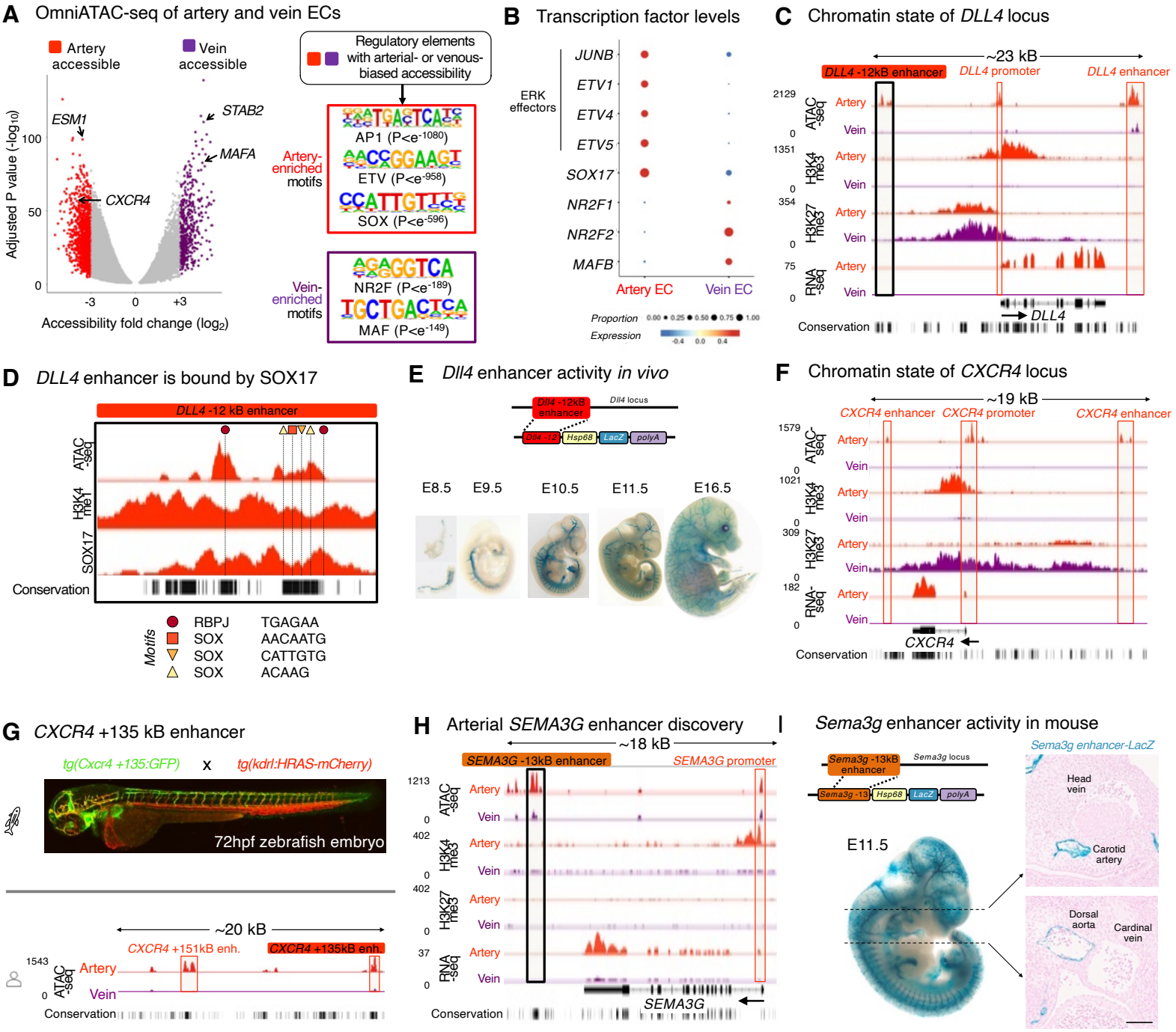


Figure 6


A OmniATAC-seq of artery and vein ECs
Artery accessible
Vein accessible
ESM1
STAB2
MAFA
CXCR4
Adjusted P value (-log10)
Accessibility fold change (log2)
Regulatory elements with arterial- or venous-biased accessibility
Artery-enriched motifs
AP1 (P<e-1080)
ETV (P<e-958)
SOX (P<e-596)
Vein-enriched motifs
NR2F (P<e-189)
MAF (P<e-149)
B Transcription factor levels
ERK effectors
JUNB
ETV1
ETV4
ETV5
SOX17
NR2F1
NR2F2
MAFB
Artery EC
Vein EC
Proportion
Expression
C Chromatin state of DLL4 locus
~23 kB
DLL4 -12kB enhancer
DLL4 promoter
DLL4 enhancer
ATAC-seq
H3K4 me3
H3K27 me3
RNA-seq
Artery
Vein
Conservation
DLL4
D DLL4 enhancer is bound by SOX17
DLL4 -12 kB enhancer
ATAC-seq
H3K4 me1
SOX17
Conservation
Motifs
RBPJ TGAGAA
SOX AACAATG
SOX CATTGTG
SOX ACAAG
E Dll4 enhancer activity in vivo
Dll4 -12kB enhancer
Dll4 locus
Dll4 -12
Hsp68
LacZ
polyA
E8.5
E9.5
E10.5
E11.5
E16.5
F Chromatin state of CXCR4 locus
~19 kB
CXCR4 enhancer
CXCR4 promoter
CXCR4 enhancer
ATAC-seq
H3K4 me3
H3K27 me3
RNA-seq
Artery
Vein
CXCR4
Conservation
G CXCR4 +135 kB enhancer
tg(Cxcr4 +135:GFP)
x
tg(kdrl:HRAS-mCherry)
72hpf zebrafish embryo
~20 kB
CXCR4 +151kB enh.
CXCR4 +135kB enh.
ATAC-seq
Artery
Vein
Conservation
H Arterial SEMA3G enhancer discovery
~18 kB
SEMA3G -13kB enhancer
SEMA3G promoter
ATAC-seq
H3K4 me3
H3K27 me3
RNA-seq
Artery
Vein
Conservation
SEMA3G
I Sema3g enhancer activity in mouse
Sema3g -13kB enhancer
Sema3g locus
Sema3g -13
Hsp68
LacZ
polyA
Sema3g enhancer-LacZ
E11.5
Head vein
Carotid artery
Dorsal aorta
Cardinal vein

892 **FIGURE LEGENDS**
893
894 **Figure 1: A roadmap for human arteriovenous differentiation reveals primed endothelial**
895 **cells**
896 A) scRNA-seq of day-0 H1 hPSCs, day-1 primitive streak, day-2 lateral mesoderm,
897 CD144+ FACS-purified day-3 artery ECs, CD144+ FACS-purified day-3 primed ECs,
898 and CD144+ FACS-purified day-4 vein ECs.
899 B) Differentially expressed genes that distinguish each hPSC-derived cell-type, as detected
900 by scRNAseq. scRNA-seq was performed on day-0 H1 hPSCs, day-1 primitive streak,
901 day-2 lateral mesoderm, CD144+ FACS-purified day-3 artery ECs, CD144+ FACS-
902 purified day-3 primed ECs, and CD144+ FACS-purified day-4 vein ECs.
903 C) scRNA-seq of day-2 lateral mesoderm, CD144+ FACS-purified day-3 artery ECs,
904 CD144+ FACS-purified day-3 primed ECs, and CD144+ FACS-purified day-4 vein ECs.
905 D) scRNA-seq of day-2 lateral mesoderm, CD144+ FACS-purified day-3 artery ECs,
906 CD144+ FACS-purified day-3 primed ECs, and CD144+ FACS-purified day-4 vein ECs.
907 Gene expression is depicted in $\log_e$ unique molecular identifier (UMI) counts.
908 E) Combined immunostaining for SOX17 protein and HCR3 *in situ* hybridization for *APLNR*
909 mRNA in the indicated H1 hPSC-derived cell-types. Scale: 200 μm.
910 F) Flow cytometry of H1 CRISPRi-expressing hPSCs, day-3 artery ECs, day-3 primed ECs,
911 and day-4 vein ECs. Day 3-4 populations were pre-gated on the CD144+ EC subset
912 before depicting marker expression.
913
914 **Figure 2: Two separable steps of vein differentiation driven by temporally dynamic**
915 **VEGF/ERK activation, followed by inhibition**
916 A) Summary of the present study.
917 B) First, H1 hPSCs were differentiated into day-2 lateral mesoderm[48]. Then, lateral
918 mesoderm was then treated with different doses of VEGF pathway modulators (VEGF
919 [0-100 ng/mL] or ERK inhibitor [ERKi; PD0325901, 500 nM]) alongside other primed EC-
920 inducing signals (TGFβ inhibitor + NOTCH inhibitor + BMP inhibitor + WNT inhibitor +
921 Vitamin C) for 24 hours. Subsequently, cells were subject to vein EC differentiation for
922 24 hours. qPCR was conducted on day 4 of hPSC differentiation. In heatmaps,
923 expression is normalized to the sample with the highest expression in either panel B or
924 C. This revealed that high VEGF for 24 hours is required during primed EC specification
925 to subsequently generate vein ECs by day 4. Statistics: Wilcoxon rank sum test. Error
926 bars: standard deviation (SD). *P<0.05.
927 C) First, H1 hPSCs were differentiated into day-3 primed ECs. Then, primed ECs were then
928 treated with different doses of VEGF pathway modulators (VEGF [0-100 ng/mL] or ERK
929 inhibitor [ERKi; PD0325901, 500 nM]) alongside other vein EC-inducing signals (TGFβ
930 inhibitor + NOTCH inhibitor + WNT agonist + Vitamin C) for 24 hours. qPCR was
931 conducted on day 4 of hPSC differentiation. In heatmaps, expression is normalized to
932 the sample with the highest expression in either panel B or C. This revealed that ERK
933 inhibition for 24 hours is required to generate day-4 vein ECs, and that after cells acquire
934 endothelial identity, VEGF/ERK is dispensable for the continued expression of pan-EC
935 markers. Statistics: Wilcoxon rank sum test. Error bars: SD. **P<0.01, *P<0.05, n.s.: not
936 significant.

D) First, H1 hPSCs were differentiated into day-3 primed ECs. Then, primed ECs were then treated with either VEGF (100 ng/mL) or ERK inhibitor (PD0325901, 500 nM) alongside other vein EC-inducing signals (TGFβ inhibitor + NOTCH inhibitor + WNT agonist + Vitamin C) for 1-24 hours, followed by qPCR. This revealed that ERK inhibition for 12 hours significantly upregulated venous marker expression.
E) Bulk-population RNA-seq of FACS-purified CD144+ day-3 primed ECs and day 4-vein ECs generated from H1 hPSCs. Differentially expressed genes are colored purple.
F) CUT&RUN profiling of H3K4me3 and H3K27me3 and bulk RNA-seq H1 hPSC-derived day-3 artery ECs, day-3 primed ECs, and day-4 vein ECs.

Figure 3: Comparison of differentiation methods suggests that temporally dynamic VEGF modulation is crucial to impart venous identity

A) scRNA-seq of hPSCs differentiated into ECs using various protocols, with the number of days of differentiation indicated. scRNA-seq datasets were subclustered at the same resolution to identify population heterogeneity. Clusters were annotated by marker expression. The proportion of ECs is indicated. scRNA-seq data were obtained from this study, Ang et al., 2022[48], Pan et al., 2024[58], McCracken et al., 2019[62], Paik et al., 2018[60], and Nikolova et al., 2025[59].
B) scRNA-seq of differentiated hPSC populations described in **Fig. 3A**, and ECs were computationally isolated. An expression module score[93] of arterial markers defined *in vivo* by Hou et al., 2022[24] is shown. Statistics: Wilcoxon rank sum test. **P<0.01.
C) scRNA-seq of differentiated hPSC populations described in **Fig. 3A**, and ECs were computationally isolated. An expression module score[93] of venous markers defined *in vivo* by Hou et al., 2022[24] is shown. Statistics: Wilcoxon rank sum test. **P<0.01.

Figure 4: *Sox17*+ *Aplnr*+ endothelial cells exist in the early embryo, and contribute to both veins and arteries

A) Whole-mount Erg and Sox17 immunostaining of E8.5 mouse embryo. Ant: anterior. Post: posterior. DA: dorsal aorta (arrowhead). VV: vitelline vein (arrowhead). Arrow: single migrating Erg+ Sox17+ ECs. Scale: 100 μm.
B) Whole-mount Erg and Sox17 immunostaining of E8.5 mouse embryo.
C) Percentage of Erg+ ECs in the E9.5 dorsal aorta (DA) or cardinal vein (CV) that express Sox17. Error bars: SD.
D) Left: whole-mount Erg immunostaining of E9.5 mouse embryo. Right: E9.5 mouse embryo sectioned and immunostained for Sox17 and Erg proteins, alongside *in situ* staining for *Aplnr* mRNA. Arrow: cardinal vein. Scale: 200 μm.
E) *sox7* and *aplnrb* staining of 27-hour post fertilization zebrafish embryos by HCR3 *in situ* hybridization. Lateral view of the mid-trunk region. Scale: 50 μm.
F) 4OHT was administered *in utero* to E9.5 *Sox17-Cre*; *Aplnr-DreER*; *RLTG* mouse embryos[66,70,71], which were then isolated at E12.5, cryosectioned, and immunostained for Erg and GFP proteins. Arrows: GFP+ vein ECs. Scale: 100 μm. Error bars: SD.

Figure 5: SOXF transcription factors are required for human vein EC specification *in vitro*

A) scRNA-seq of day-2 lateral mesoderm, CD144+ FACS-purified day-3 artery ECs, CD144+ FACS-purified day-3 primed ECs, and CD144+ FACS-purified day-4 vein ECs

generated from H1 hPSCs. Arrows indicate that primed ECs express *SOX7*, *SOX17*, and *SOX18*.

B) H1 CRISPRi-expressing hPSCs were transduced with sgRNAs targeting *SOX7*, *SOX17*, and *SOX18* (*SOXF* triple knockdown [TKD]), and then subsequently differentiated into artery and vein ECs.

C) scRNA-seq of day-3 artery ECs and day-4 vein ECs generated from H1 control vs. *SOXF* TKD hPSCs. Mesenchymal cells were computationally excluded.

D) scRNA-seq of day-3 artery ECs, day-3 primed ECs, and day-4 vein ECs generated from H1 control vs. *SOXF* TKD hPSCs. Mesenchymal cells were computationally excluded. A transcriptional module score[93] computed from a panel of *in vivo*-defined arterial marker genes *in vivo*[24] is shown. Statistics: Wilcoxon rank sum test. **P<0.01.

E) scRNA-seq of day-3 artery ECs and day-4 vein ECs generated from H1 control vs. *SOXF* TKD hPSCs. Mesenchymal cells were computationally excluded.

F) scRNA-seq of day-3 artery ECs, day-3 primed ECs, and day-4 vein ECs generated from H1 control vs. *SOXF* TKD hPSCs. Mesenchymal cells were computationally excluded. A transcriptional module score[93] computed from a panel of *in vivo*-defined venous marker genes *in vivo*[24] is shown. Statistics: Wilcoxon rank sum test. **P<0.01.

G) scRNA-seq of day-3 artery ECs and day-4 vein ECs generated from H1 control vs. *SOXF* TKD hPSCs. Mesenchymal cells were computationally excluded. Differentially expressed genes between control vs. *SOXF* TKD ECs are colored.

H) OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 primed ECs and day-4 vein ECs.

Figure 6: Chromatin analysis links VEGF/ERK inhibition to vein EC specification

A) OmniATAC-seq of H1 hPSC-derived day-3 artery ECs and day-4 vein ECs. Left: genetic loci preferentially accessible in either artery or vein ECs are colored. Right: transcription factor motifs enriched in artery- or vein-accessible chromatin elements.

B) Expression of AP1, ETV, SOX, NR2F, and MAF family transcription factors at the mRNA level, as shown by scRNAseq of FACS-purified CD144+ DLL4+ CD73lo/- day-3 artery ECs and CD144+ DLL4- CD73hi day-4 vein ECs generated from H1 hPSCs.

C) OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 artery ECs and day-4 vein ECs.

D) OmniATAC-seq and CUT&RUN analysis of the *DLL4* -12kb enhancer[78] in H1 hPSC-derived day-3 artery ECs, with transcription factor motifs labeled.

E) LacZ staining of E8.5-E16.5 mouse embryos bearing a *Dll4* -12 kB enhancer transgene driving *LacZ* expression[78].

F) OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 artery ECs and day-4 vein ECs.

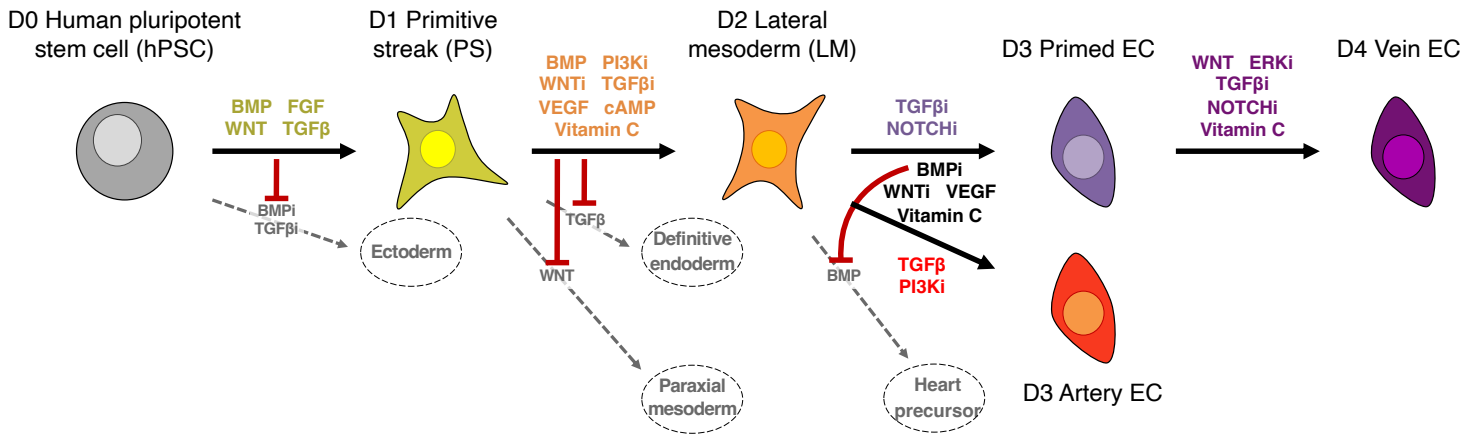
G) *Top*: image of a 3 day-post-fertilization (dpf) zebrafish bearing a *Cxcr4* +135 kB enhancer driving *GFP* expression, together with a *kdrl:HRAS-mCherry* transgene to label ECs[79]. *Bottom*: ATAC-seq of H1 hPSC-derived day-3 artery ECs and day-4 vein ECs, highlighting the *CXCR4* +135 kB enhancer whose ortholog was tested in the zebrafish transgenic assay[79].

H) OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 artery ECs and day-4 vein ECs.

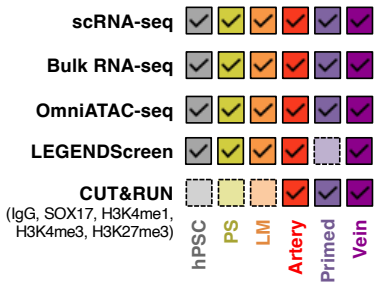
1028 l) LacZ staining of E11.5 mouse embryo bearing a *Sema3g* -13 kB enhancer transgene
1029 driving *LacZ* expression (VISTA Enhancer Browser ID hs2179)[82,83]. Scale: 100 μm.

Supplemental Figure 1

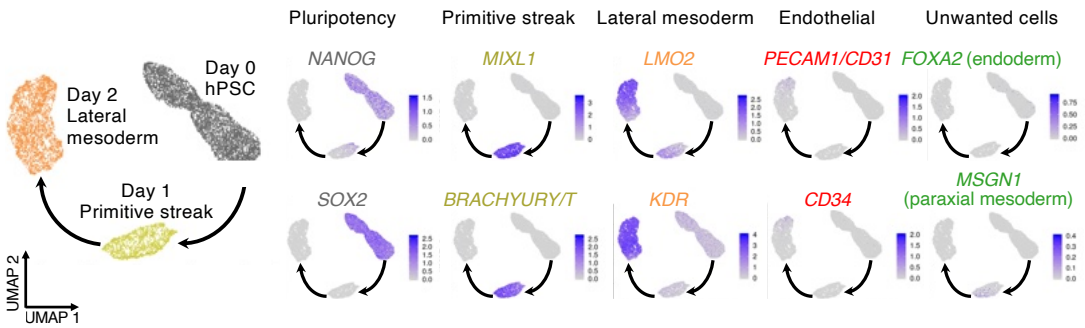
A Differentiation of hPSCs into artery and vein ECs



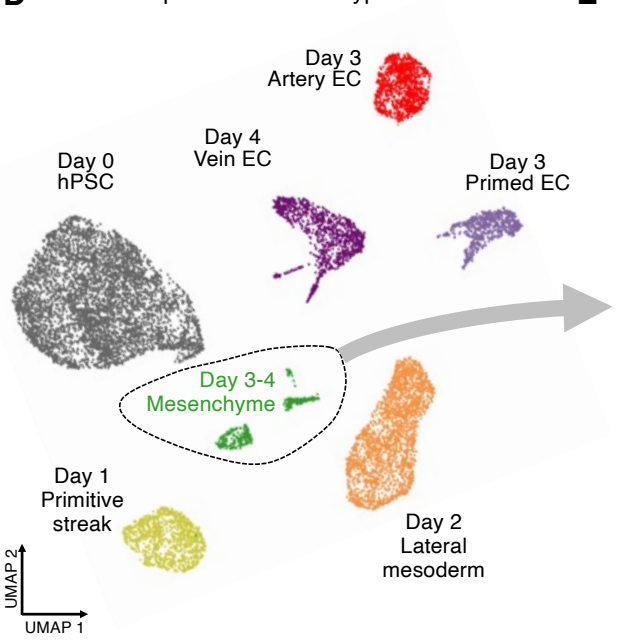
B Genomic characterizations



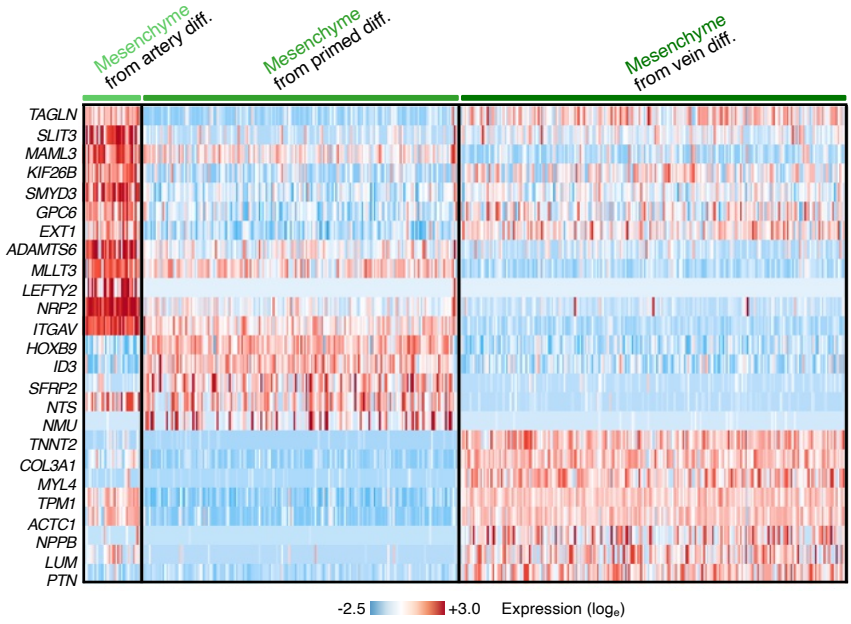
C Stepwise differentiation of hPSCs into primitive streak and lateral mesoderm



D scRNA-seq of unsorted cell-types

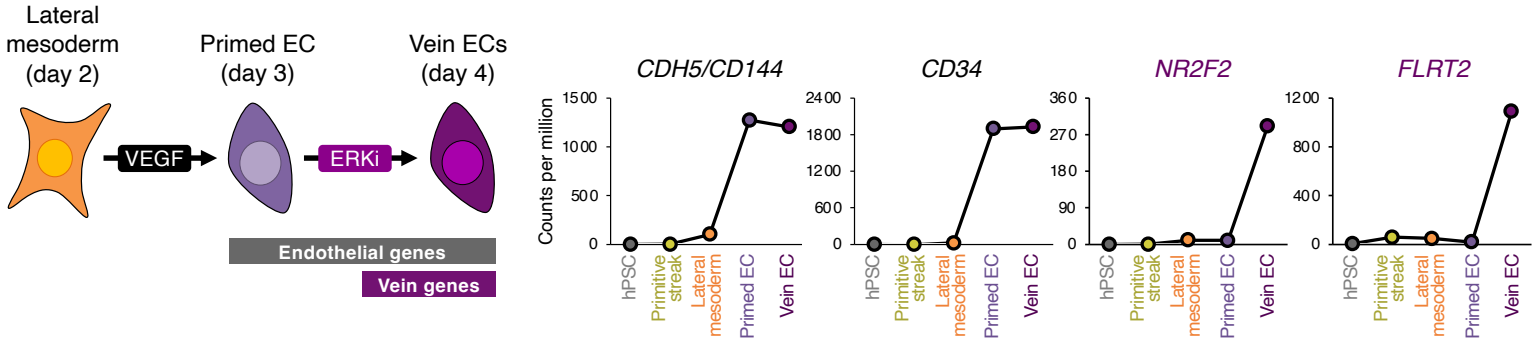


E Diverse mesenchymal cells arising on days 3-4 of differentiation

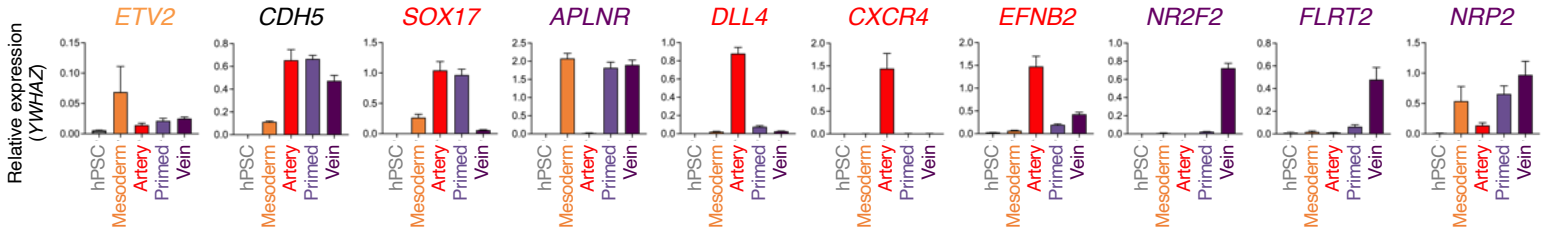


Supplemental Figure 2

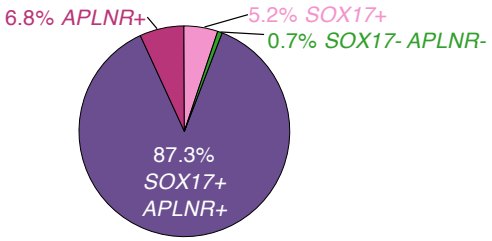
A Bulk RNA-seq of vein differentiation



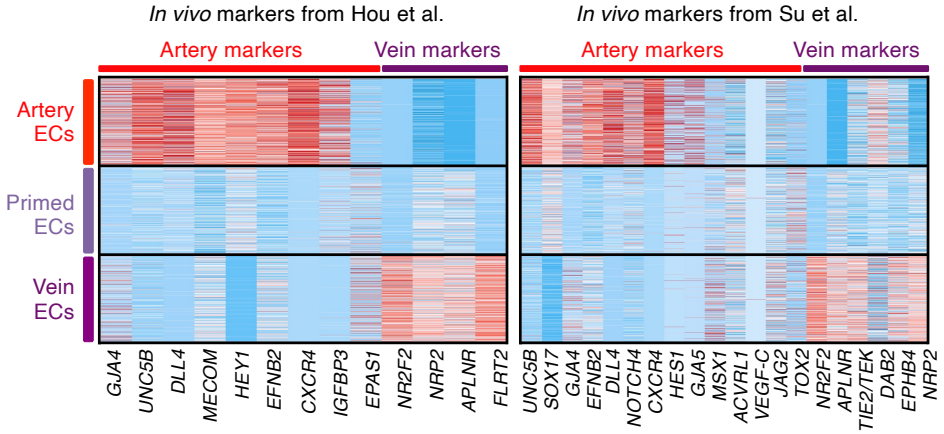
B qPCR validation of scRNA-seq



C Primed ECs co-express *SOX17* and *APLNR*

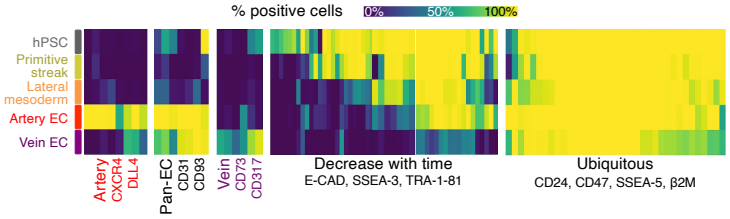


D Expression of arteriovenous markers by hPSC-derived ECs

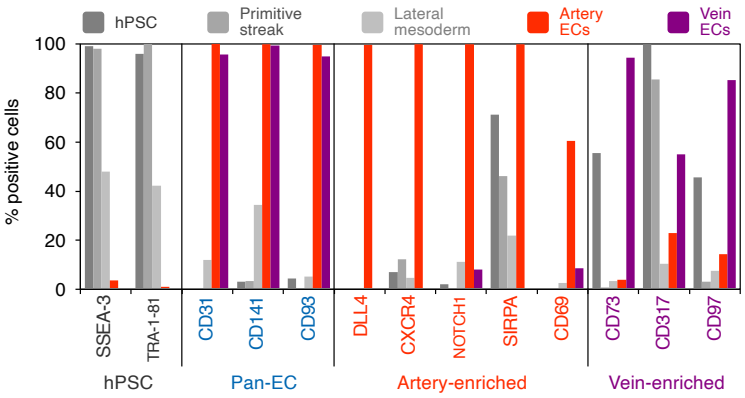


Supplemental Figure 3

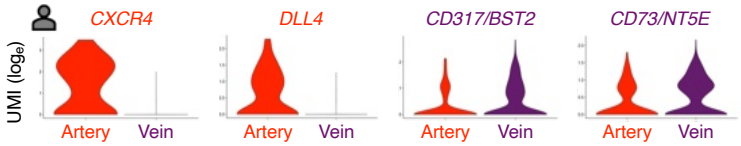
A LEGENDScreen flow cytometry of cell-surface markers



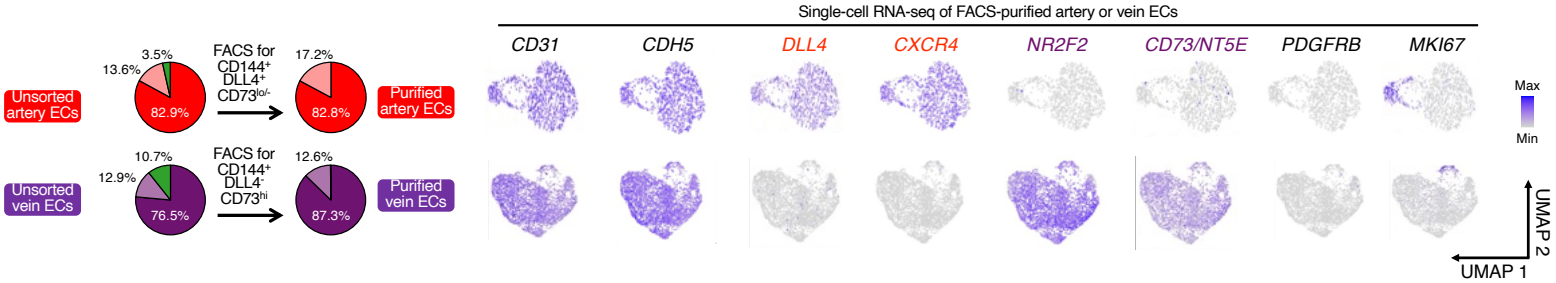
B High-throughput screening of cell-surface markers



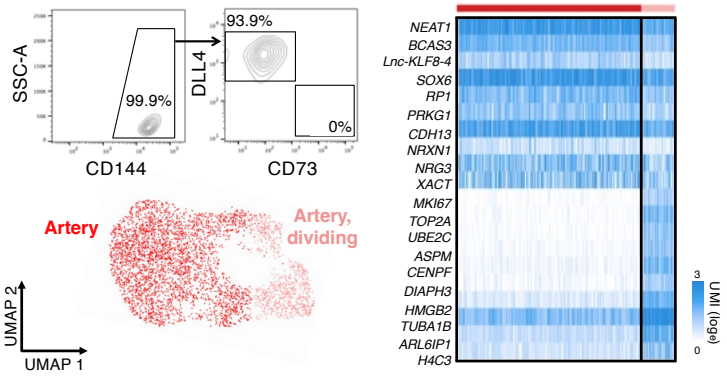
C CS17 human fetal ECs (scRNA-seq; Calvanese et al.)



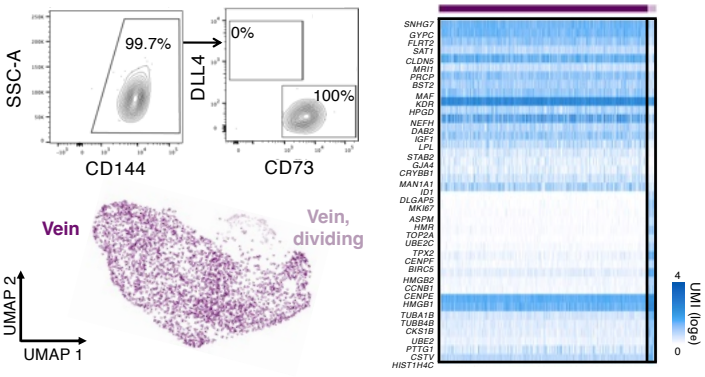
D scRNA-seq of FACS-purified CD144+ artery and vein ECs



E scRNA-seq of FACS-purified CD144+ artery ECs

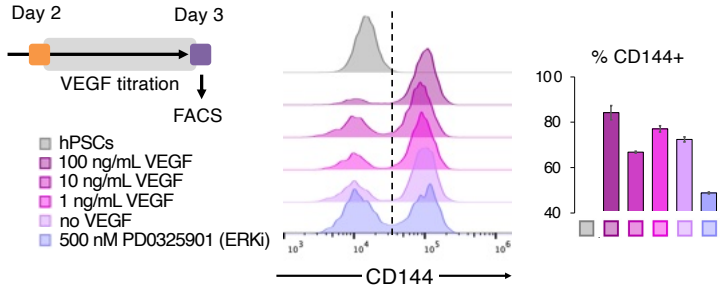


F scRNA-seq of FACS-purified CD144+ vein ECs

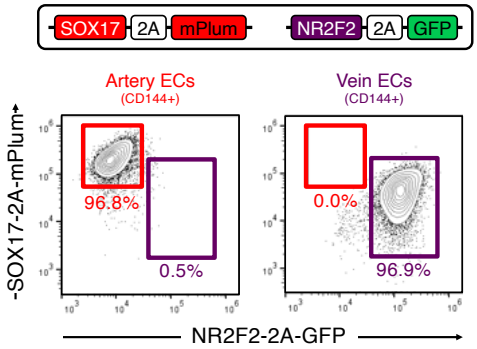


Supplemental Fig. 4

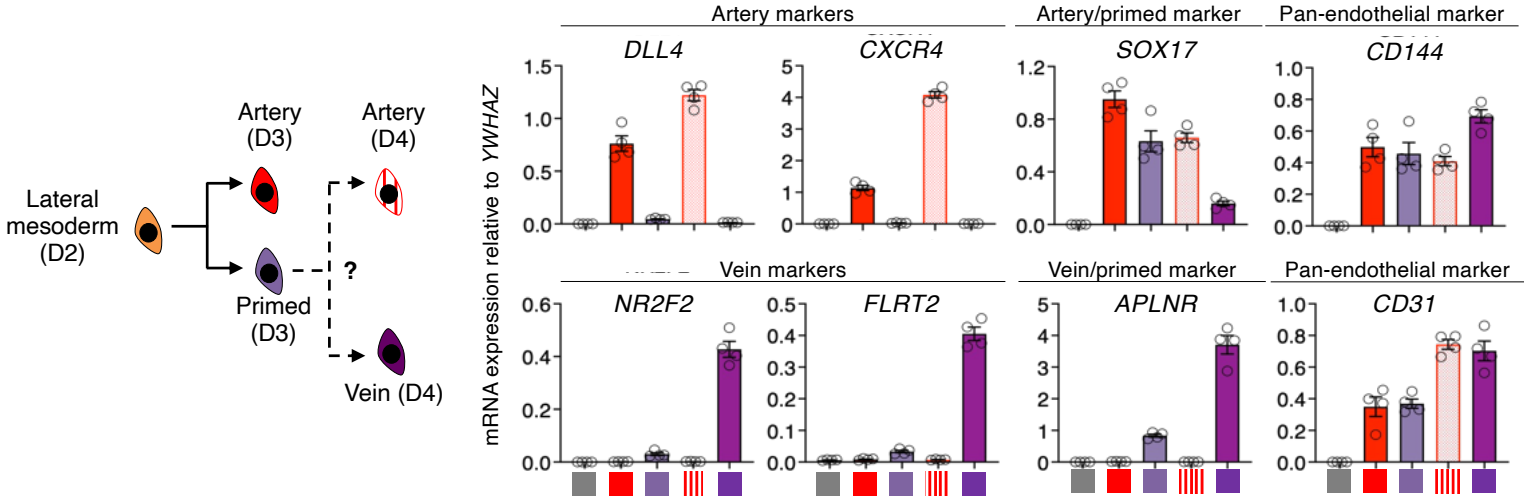
A VEGF is required during primed induction to acquire EC identity



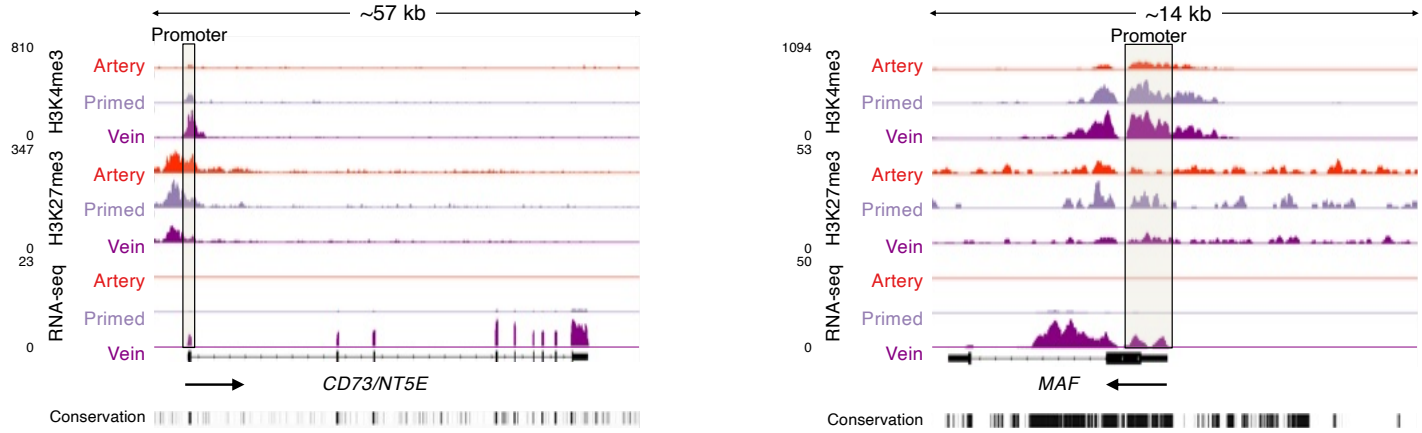
B VEGF/ERK inhibition induces NR2F2+ vein ECs



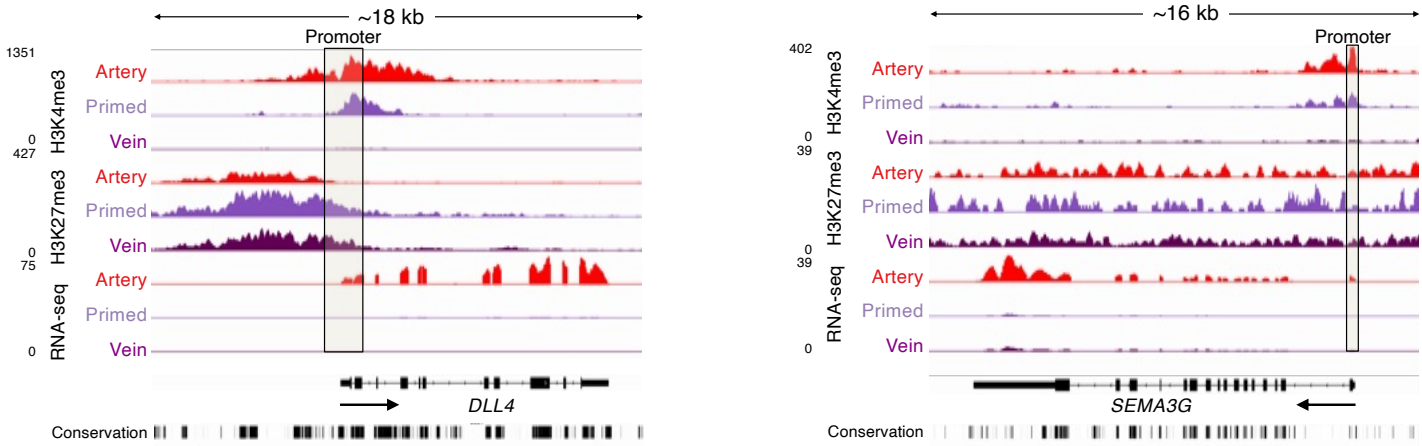
C Primed ECs can generate both artery and vein ECs *in vitro*



D Promoters of future venous genes are bivalently marked in primed ECs

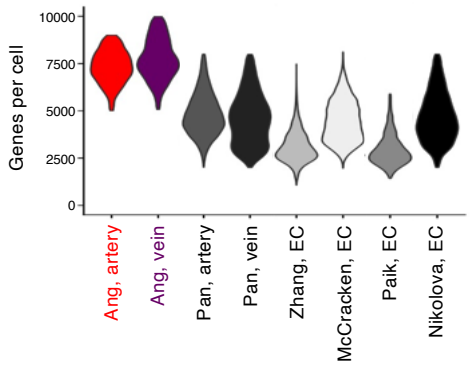


E Promoters of future arterial genes are bivalently marked in primed ECs

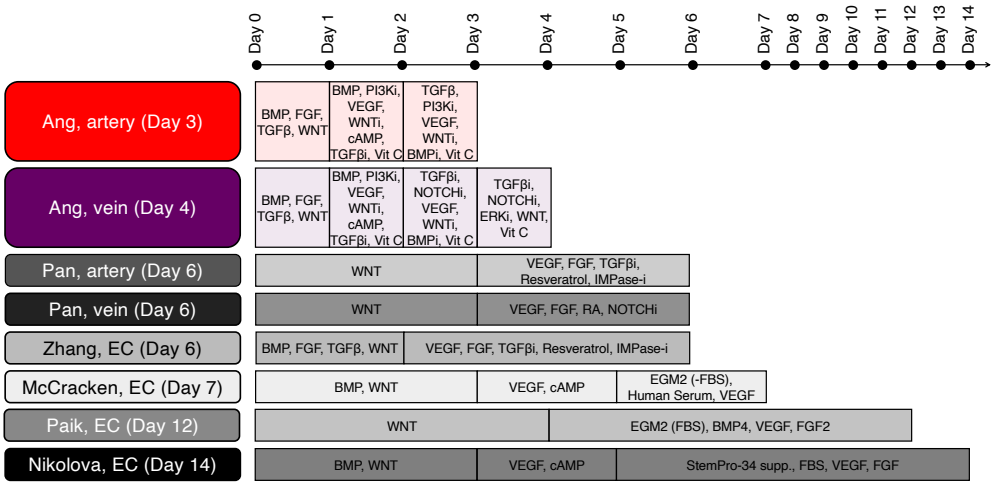


Supplemental Figure 5

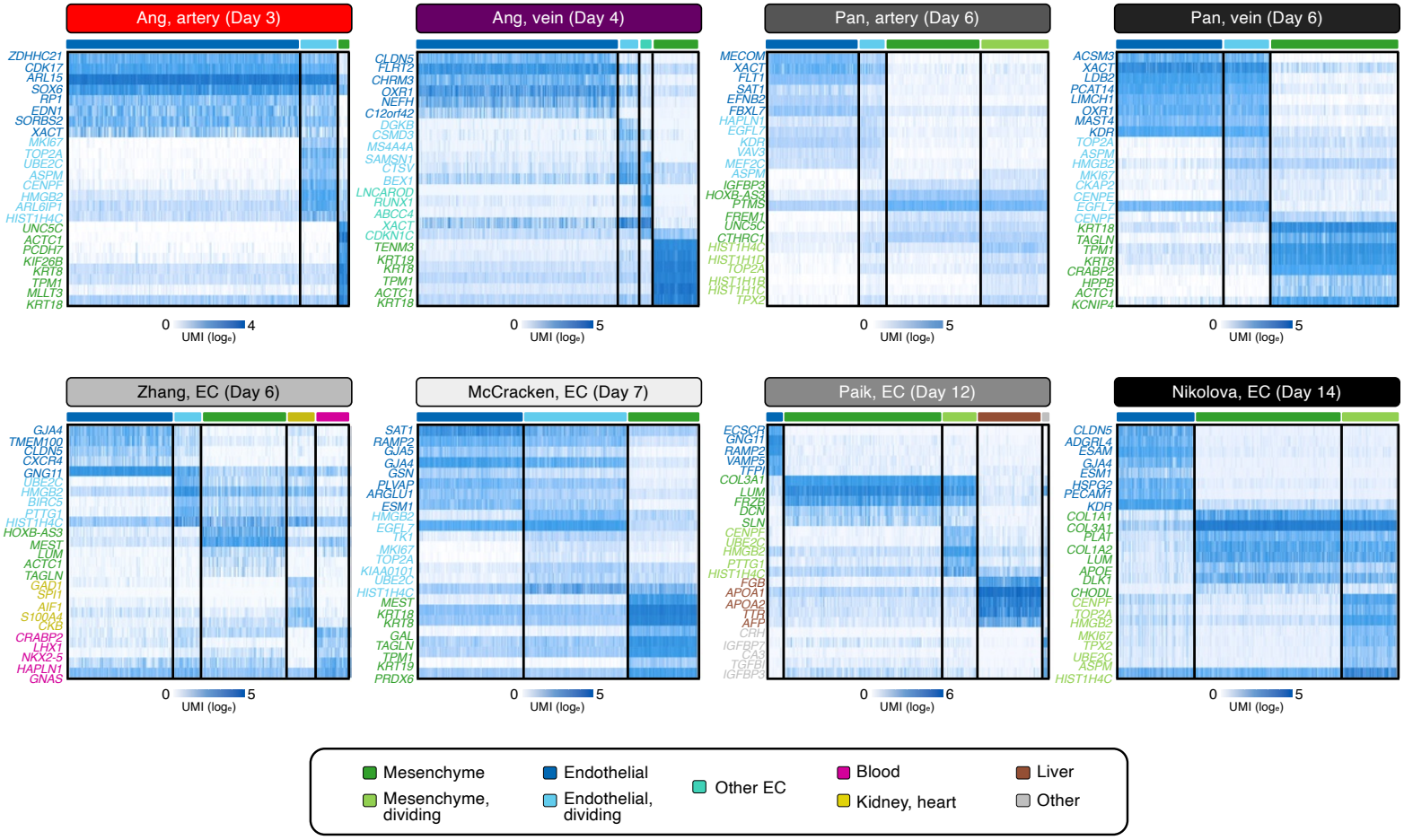
A Quality control of scRNA-seq datasets



B Summary of EC differentiation methods

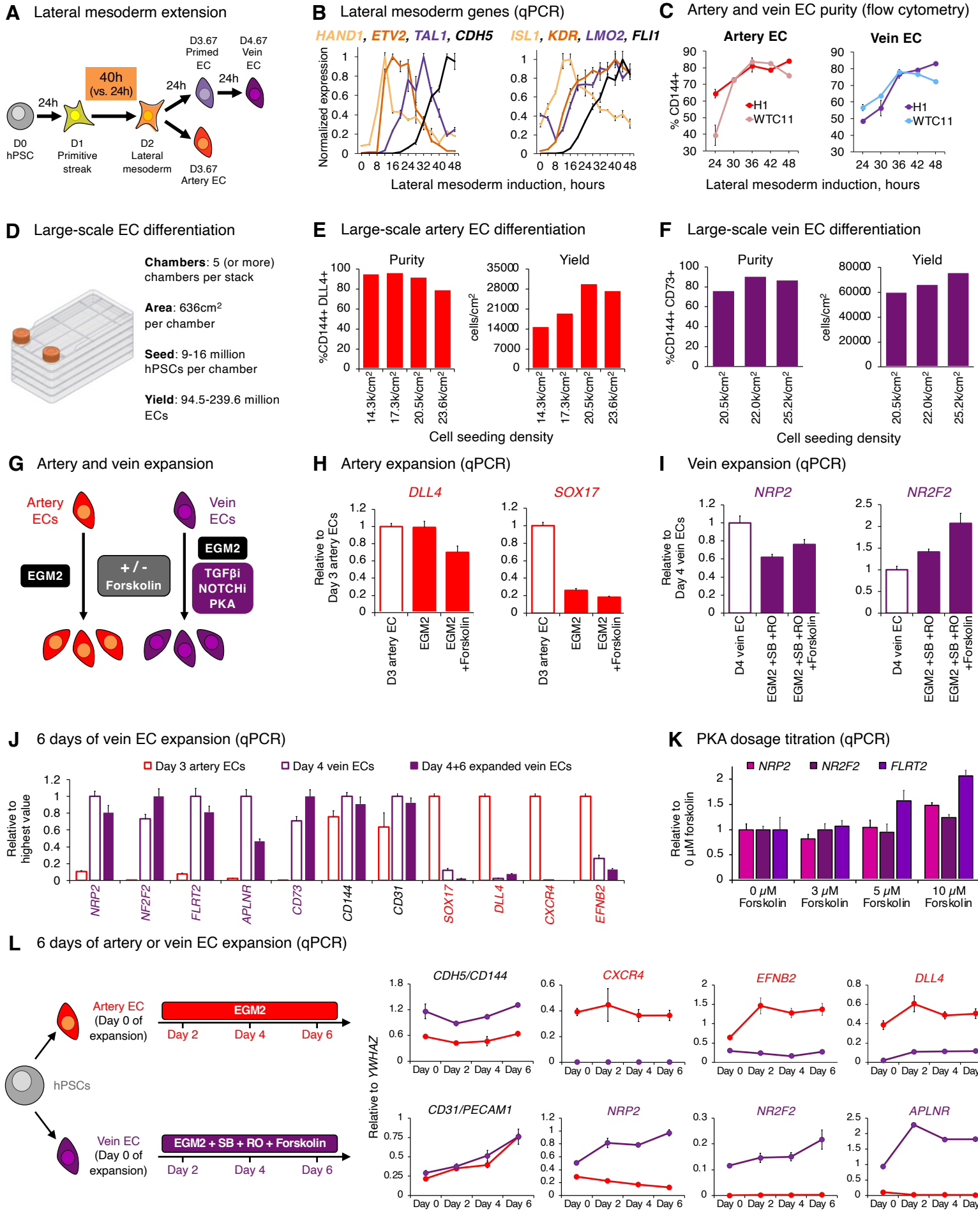


C Cellular diversity resulting from each differentiation protocol (scRNA-seq)



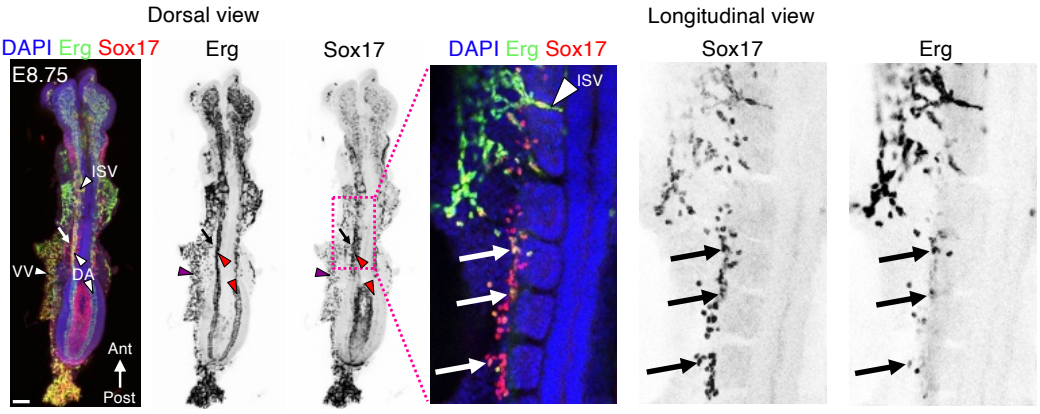
Supplemental Figure 6

Supplemental Figure 7

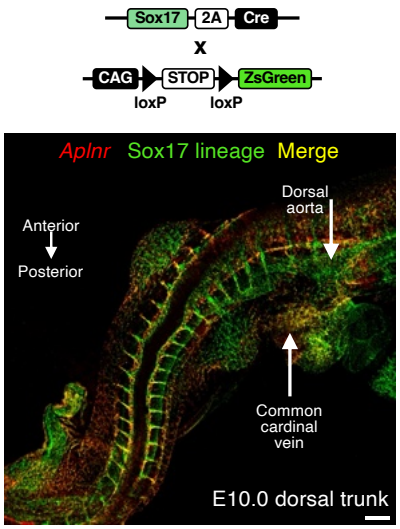


Supplemental Figure 8

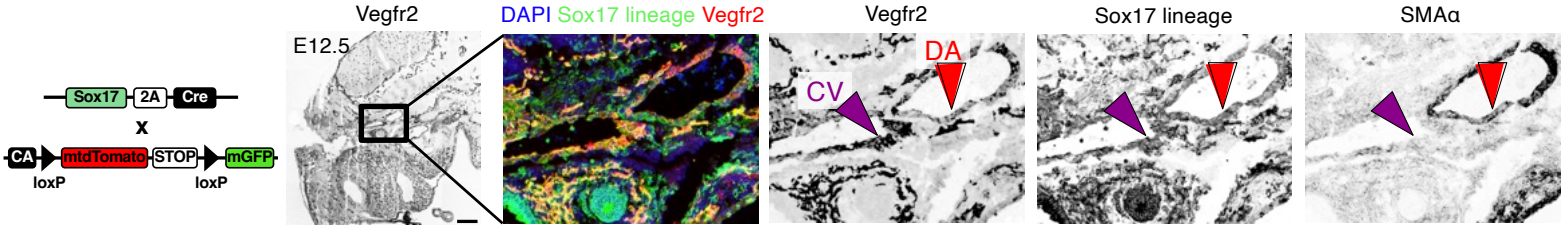
A Widespread Sox17 expression in E8.75 mouse embryo ECs



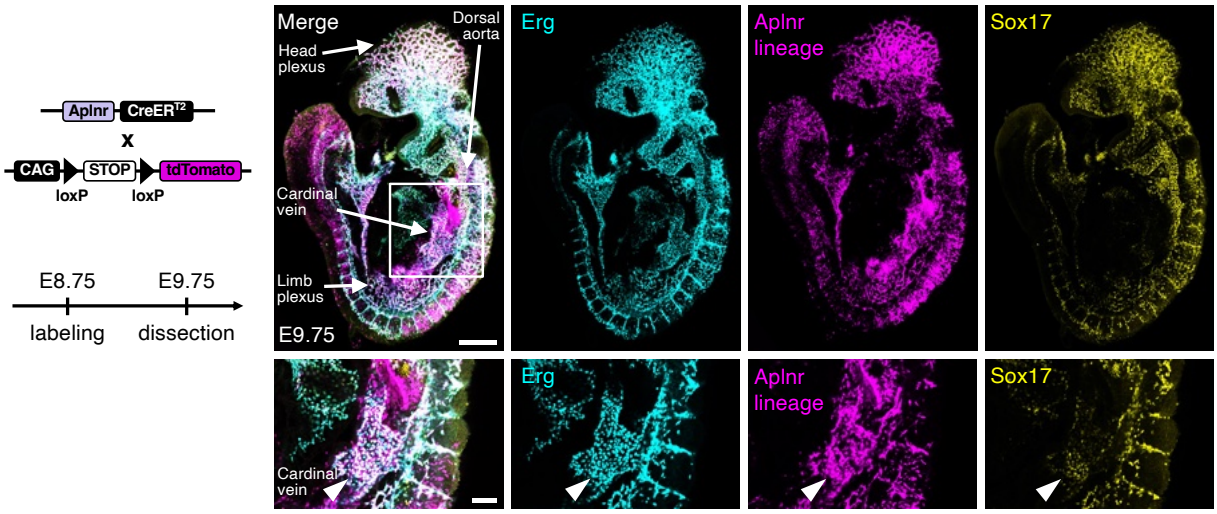
B *Sox17-Cre* labels artery and vein ECs



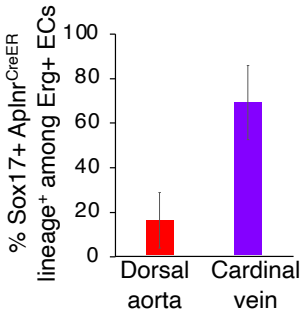
C *Sox17-Cre* lineage tracing of E12.5 mouse embryo



D *Aplnr-CreER* labeling (E9.75 mouse embryo)

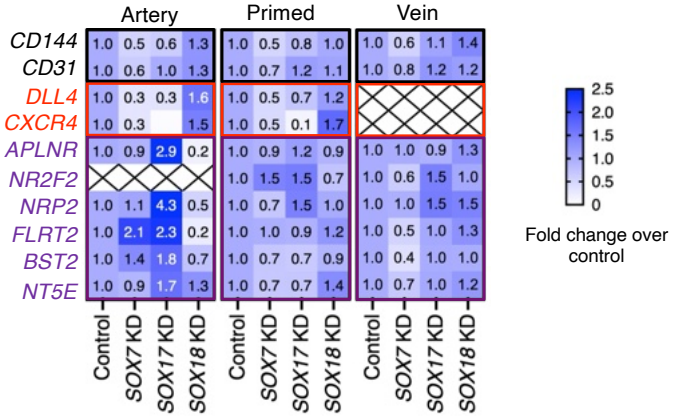


E Co-expression of Sox17 and *Aplnr-CreER* lineage in E9.5 mouse embryos

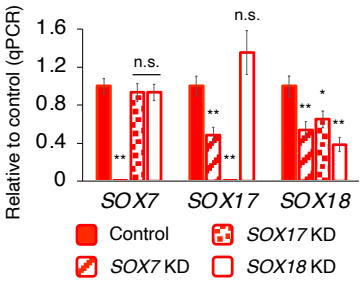


Supplemental Figure 9

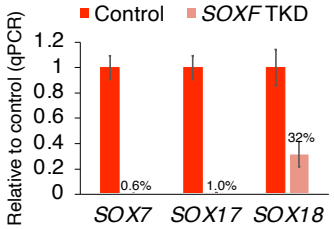
A Single knockdown of *SOX7*, *SOX17*, or *SOX18* (qPCR)



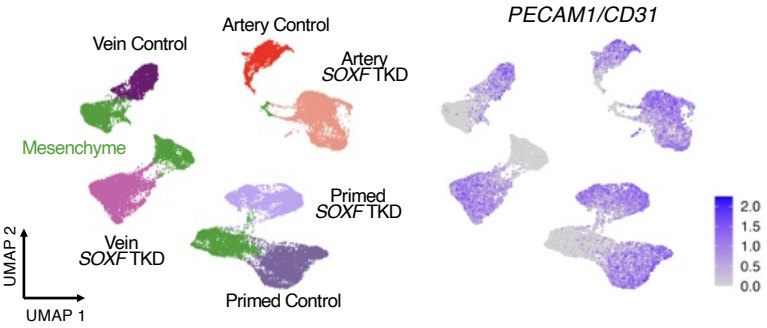
B Interdependency of *SOXF* factors in artery ECs (qPCR)



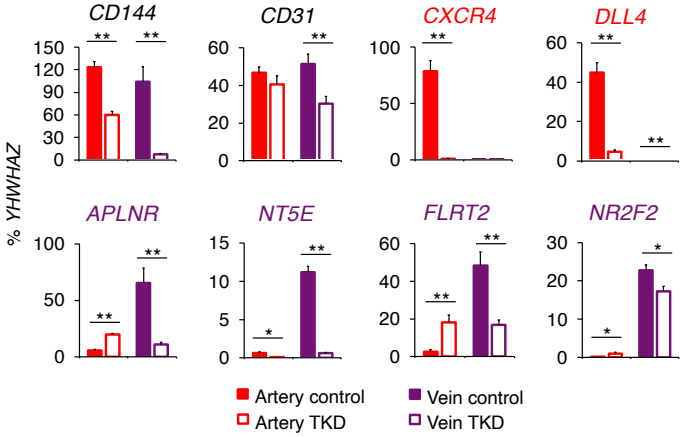
C *SOXF* TKD in artery ECs



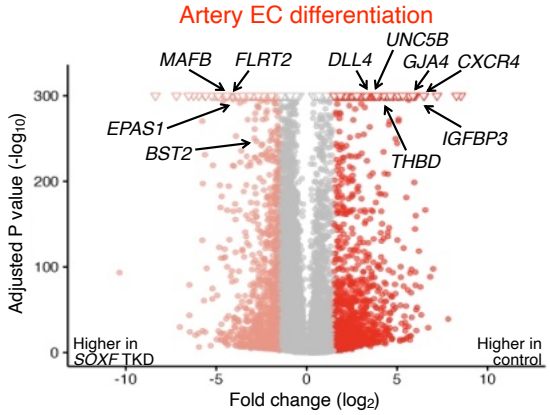
D Control and *SOXF* TKD differentiated cells (scRNA-seq)



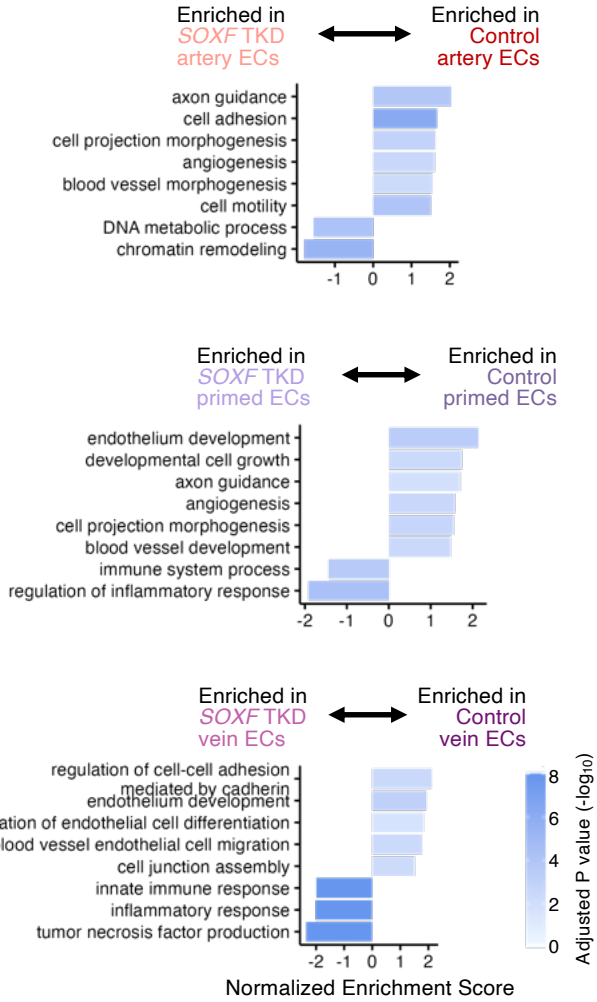
E Gene expression changes in *SOXF* TKD artery and vein ECs (qPCR)



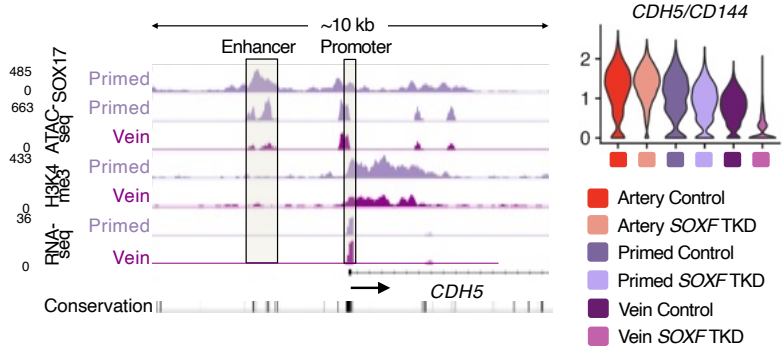
F *SOXF* depletion abrogates artery EC differentiation



G Gene ontology analysis of *SOXF*-deficient ECs (scRNA-seq)

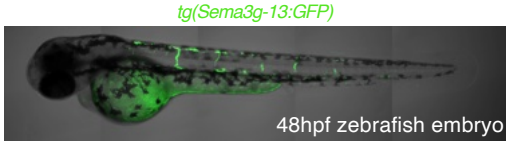


H SOX17 binds to the *VE-CADHERIN* gene in primed ECs

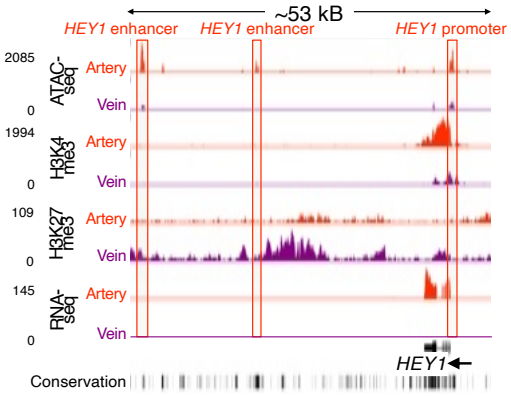


Supplemental Figure 10

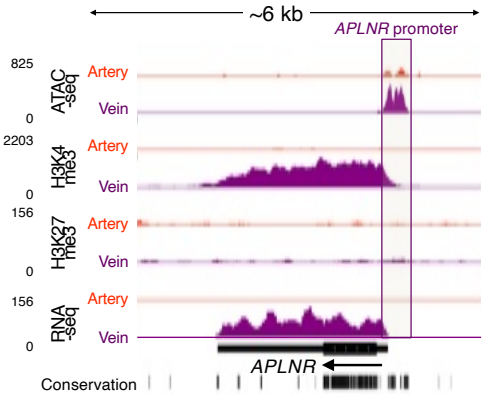
A *Sema3g* enhancer activity in zebrafish



B Arterial *HEY1* enhancer discovery



C Chromatin state of *APLNR* locus



1030 **SUPPLEMENTAL FIGURE LEGENDS**
1031
1032 **Figure S1: Cellular diversity during early steps of hPSC differentiation**
1033 A) Summary of hPSC differentiation approach[48]. i: inhibition.
1034 B) Summary of gene expression, chromatin state, and cell-surface marker profiling
1035 conducted as part of this study. Dotted lines indicate that profiling was not conducted on
1036 a given cell-type.
1037 C) Gene expression on days 0, 1, and 2 of H1 hPSC differentiation, as detected by
1038 scRNAseq.
1039 D) scRNAseq of day-0 H1 hPSCs that were differentiated into day-1 primitive streak, day-
1040 2 lateral mesoderm, day-3 artery ECs, day-3 primed ECs, and day-4 vein ECs. None of
1041 cell populations shown here were FACS purified and consequently, mesenchymal cells
1042 were detected alongside artery, primed, and vein ECs on days 3-4 of differentiation. The
1043 same day-0 hPSC, day-1 primitive streak, and day-2 primitive streak scRNAseq
1044 datasets were also used in **Fig. 1A**; however, FACS-purified downstream EC
1045 populations are shown in **Fig. 1A**.
1046 E) scRNAseq of H1 hPSC-derived mesenchymal cells that arose alongside day-3 artery
1047 ECs, day-3 primed ECs, and day-4 vein ECs. Differentially expressed genes that
1048 distinguish these three different types of mesenchyme are shown. Diff: differentiation.

1049 **Figure S2: Characterization of primed endothelial cells *in vitro***
1050 A) Bulk RNA-seq of the indicated H1 hPSC-derived cell-types. FACS was used to purify
1051 CD144+ primed and vein ECs for RNA-seq. i: inhibition.
1052 B) qPCR of the indicated H1 hPSC-derived cell-types. qPCR data normalized to reference
1053 gene *YWHAZ* (i.e., *YWHAZ* levels = 1.0). Error bars: standard deviation (SD).
1054 C) scRNAseq of H1 hPSC-derived day-3 primed EC populations, showing the percentage
1055 of cells that expressed *APLNR* and/or *SOX17* mRNAs.
1056 D) scRNAseq of FACS-purified CD144+ DLL4+ CD73lo/- day-3 artery ECs, CD144+ day-
1057 3 primed ECs, and CD144+ DLL4- CD73hi day-4 vein ECs generated from H1 hPSCs.
1058 In each of these *in vitro* cell-types, expression of *in vivo* arterial and venous markers
1059 reported by Hou et al., 2022[24] and Su et al., 2018[23] is shown.
1060
1061 **Figure S3: Cell-surface markers of human arteriovenous identity**
1062 A) LEGENDScreen high-throughput flow cytometry of cell-surface markers in day-0
1063 hPSCs, day-1 primitive streak, day-2 lateral mesoderm, day-3 artery ECs, and day-4
1064 vein ECs. Day-3 artery and day-4 vein EC populations were pre-gated on the CD144+
1065 EC subset before depicting marker expression.
1066 B) LEGENDScreen high-throughput flow cytometry of cell-surface markers in day-0
1067 hPSCs, day-1 primitive streak, day-2 lateral mesoderm, day-3 artery ECs, and day-4
1068 vein ECs. Day-3 artery and day-4 vein EC populations were pre-gated on the CD144+
1069 EC subset before depicting marker expression.
1070 C) scRNAseq of human Carnegie Stage 17 (CS17) fetal ECs. scRNAseq data were
1071 obtained from Calvanese et al., 2022[94].
1072 D) Left: Population heterogeneity of H1 hPSC-derived day-3 artery ECs and day-4 vein
1073 ECs, before and after FACS purification based on the indicated cell-surface marker
1074 combinations. Proportions of cell clusters in scRNAseq data are shown. Right:

scRNAseq of FACS-purified CD144+ DLL4+ CD73lo/- day-3 artery ECs and CD144+ DLL4- CD73hi day-4 vein ECs generated from H1 hPSCs.

E) scRNAseq of FACS-purified CD144+ DLL4+ CD73lo/- day-3 artery ECs generated from H1 hPSCs. Left: FACS isolation strategy. Right: subclustering was performed to assess any potential population heterogeneity, and differentially expressed genes that distinguished cell subsets are shown.

F) scRNAseq of FACS-purified CD144+ DLL4- CD73hi day-4 vein ECs generated from H1 hPSCs. Left: FACS isolation strategy. Right: subclustering was performed to assess any potential population heterogeneity, and differentially expressed genes that distinguished cell subsets are shown.

Figure S4: Generation and characterization of primed endothelial cells *in vitro*

A) First, H1 hPSCs were differentiated into day-2 lateral mesoderm[48]. Then, lateral mesoderm was then treated with different doses of VEGF pathway modulators (VEGF [0-100 ng/mL] or ERK inhibitor [ERKi; PD0325901, 500 nM]) alongside other primed EC-inducing signals (TGFβ inhibitor + NOTCH inhibitor + BMP inhibitor + WNT inhibitor + Vitamin C) for 24 hours. Flow cytometry was conducted on day 3 of hPSC differentiation. This revealed that high VEGF for 24 hours is required during primed EC specification to efficiently generate ECs by day 3.

B) Flow cytometry of H1 *SOX17-2A-mPlum; NR2F2-2A-GFP* hPSCs differentiated into day-3 artery ECs or day-5 vein ECs. Day-4 vein ECs were maintained in the same vein induction media until day 5. Day 3-5 populations were pre-gated on the CD144+ EC subset before depicting marker expression.

C) First, H1 hPSCs were differentiated into day-3 primed ECs. Then, primed ECs were then treated with different doses of VEGF pathway modulators (VEGF [0-100 ng/mL] or ERK inhibitor [ERKi; PD0325901, 500 nM]) alongside other vein EC-inducing signals (TGFβ inhibitor + NOTCH inhibitor + WNT agonist + Vitamin C) for 1-24 hours. qPCR was conducted on day 4 of hPSC differentiation, and expression is normalized to the sample with the highest expression. This revealed that ERK inhibition for 12 hours significantly upregulated venous marker expression.

D) CUT&RUN profiling of H3K4me3 and H3K27me3 and bulk RNA-seq H1 hPSC-derived day-3 artery ECs, day-3 primed ECs, and day-4 vein ECs.

E) CUT&RUN profiling of H3K4me3 and H3K27me3 and bulk RNA-seq H1 hPSC-derived day-3 artery ECs, day-3 primed ECs, and day-4 vein ECs.

Figure S5: Comparison of methods to differentiate hPSCs into ECs

A) scRNA-seq of hPSCs differentiated into ECs using various protocols, depicting the number of genes detected per single cell as a quality control metric.

B) Summary of protocols to differentiate hPSCs into ECs[48,58-62] used to generate the scRNA-seq datasets described in **Fig. 3A**.

C) scRNA-seq of differentiated hPSC populations described in **Fig. 3A**. Subclustering was performed to assess population heterogeneity, and differentially expressed genes that distinguished cell subsets are shown. Clusters are annotated and colored as described in **Fig. 3A**.

Figure S6: Comparison of methods to differentiate hPSCs into ECs

A) scRNA-seq of differentiated hPSC populations described in **Fig. 3A**, depicting expression of pan-endothelial, arterial, venous, and mesenchymal marker genes. Clusters are annotated and colored as described in **Fig. 3A**.

Figure S7: Improved generation and expansion of artery and vein ECs, while preserving arteriovenous identity

A) Schematic of differentiation protocol, which includes prolongation of lateral mesoderm induction from 24 hours (V1 protocol) to 40 hours (V2 protocol).
B) qPCR of lateral mesoderm and vascular marker expression in H1 hPSC-derived day-1 primitive streak cells challenged with lateral mesoderm induction media for up to 48 hours. 0 hours corresponds to day-1 primitive streak cells, prior to addition of lateral mesoderm induction media. Cells were lysed every 4 hours and qPCR was performed. For each gene, expression is normalized to the sample with the highest value, which was set to “1.0”. Error bars: SEM.
C) Effect of lateral mesoderm induction duration on the subsequent production of CD144+ artery and vein ECs from the H1 and WTC11 hPSC lines, as measured by flow cytometry. Duration of lateral mesoderm induction was varied from 24 hours (V1 protocol) to 48 hours, with all other parameters of the differentiation protocol unchanged. Error bars: SEM.
D) Schematic of CellSTACK system for large-scale differentiation of hPSCs into ECs.
E) Yield and percentage of CD144+ DLL4+ day-3.67 artery ECs generated in the large-scale CellSTACK differentiation system, depending on initial hPSC seeding density. Yield and differentiation purity were respectively measured by flow cytometry and cell counting.
F) Yield and percentage of CD73+ DLL4+ day-4.67 vein ECs generated in the large-scale CellSTACK differentiation system, depending on initial hPSC seeding density. Yield and differentiation purity were respectively measured by flow cytometry and cell counting.
G) Strategy for expanding hPSC-derived artery and vein ECs. Artery ECs are cultured in EGM2 medium[48], and vein ECs are cultured in EGM2 medium + TGFβ inhibitor + NOTCH inhibitor + PKA activator. i: inhibitor.
H) qPCR of H1 hPSC-derived artery ECs that were thawed in EGM2 medium[48] and cultured for 6 days. ROCK inhibitor (Thiazovivin, 2 μM) was added for the first day of thawing to improve cell survival, and was subsequently removed in later days. Forskolin (3 μM) was added as indicated. For each gene, expression is normalized to the sample with the highest value, which was set to “1.0”. Error bars: SEM.
I) qPCR of H1 hPSC-derived vein ECs that were thawed in EGM2 medium + SB505124 (2 μM, TGFβ inhibitor) + RO4929097 (1 μM, NOTCH inhibitor)[48] and cultured for 6 days. ROCK inhibitor (Thiazovivin, 2 μM) was added for the first day of thawing to improve cell survival, and was subsequently removed in later days. Forskolin (3 μM) was added as indicated. For each gene, expression is normalized to the sample with the highest value, which was set to “1.0”. Error bars: SEM.
J) qPCR of H1 hPSC-derived vein ECs that were thawed in EGM2 medium + SB505124 (2 μM, TGFβ inhibitor) + RO4929097 (1 μM, NOTCH inhibitor) + Forskolin (10 μM, adenylate cyclase agonist) and cultured for 6 days. ROCK inhibitor (Thiazovivin, 2 μM) was added for the first day of thawing to improve cell survival, and was subsequently removed in later days. As a control, H1 hPSC-derived artery ECs prior to expansion

were also analyzed. qPCR data are normalized to the sample with the highest expression, which was set to “1.0”. Error bars: SEM.

K) qPCR of H1 hPSC-derived vein ECs that were thawed in EGM2 medium + SB505124 (2 μM, TGFβ inhibitor) + RO4929097 (1 μM, NOTCH inhibitor), in the presence or absence of Forskolin (3-10 μM, adenylate cyclase agonist) and cultured for 6 days. qPCR data are normalized to the sample lacking forskolin, which was set to “1.0”. Error bars: SEM.

L) qPCR of pan-endothelial, arterial, and venous marker expression over the course of 6-day expansion of SUN004.2 *CAG-mScarlet* hPSC-derived artery and vein ECs. Artery ECs were expanded in EGM2 medium, whereas vein ECs were expanded in EGM2 medium + SB505124 (2 μM, TGFβ inhibitor) + RO4929097 (1 μM, NOTCH inhibitor) + Forskolin (10 μM, adenylate cyclase agonist). qPCR data normalized to reference gene *YWHAZ* (i.e., *YWHAZ* levels = 1.0). Error bars: SEM.

Figure S8: Existence and fate of *Sox17*+ *Aplnr*+ endothelial cells in the mouse embryo

A) Whole-mount Erg and Sox17 immunostaining of E8.75 mouse embryo. DA: dorsal aorta (arrowhead). VV: vitelline vein (arrowhead). ISV: intersomitic vessel (arrowhead). Arrows: solitary Sox17+ Erg+ ECs. Ant: anterior. Post: posterior. Scale: 100 μm.

B) Whole-mount *Aplnr* mRNA staining of E10 *Sox17*^{Cre}*; R26*^{zsGreen} mouse embryos[66,95] by HCR3 *in situ* hybridization. Scale: 100 μm.

C) E12.5 *Sox17*^{Cre}*; R26*^{mTmG} mouse embryos[66,96] were sectioned and stained for Vegfr2, SMAα, and GFP proteins. Scale: 200 μm.

D) 4-hydroxytamoxifen (4OHT) was administered *in utero* to E8.75 *Aplnr-CreER*; *R26-tdTomato* mouse embryos[20,95], which were then isolated at E9.75 and immunostained for Sox17 and Erg proteins. Bottom row arrows: cardinal vein. Scale: 200 μm (top row), 50 μm (bottom row).

E) 4-hydroxytamoxifen (4OHT) was administered *in utero* to E8.75 *Aplnr-CreER*; *R26-tdTomato* mouse embryos[20,95], which were then isolated at E9.75 and immunostained for Sox17 and Erg proteins. Quantification of Erg+ ECs in the dorsal aorta and cardinal vein that co-expressed tdTomato (indicative of *Aplnr-CreER* activity) and Sox17. Error bars: SD. **P<0.01.

Figure S9: Triple CRISPRi knockdown of *SOXF* transcription factors impairs both human artery and vein EC differentiation *in vitro*

A) qPCR of day-3 artery ECs derived from control and single *SOX7*-, *SOX17*-, or *SOX18*-knockdown H1 CRISPRi hPSC lines. Gene expression normalized to levels in control hPSC-derived artery ECs, which was set as “1.0”. “X” indicates that qPCR data were not shown, because gene expression was under 2% of *YWHAZ* in control samples. Statistics: unpaired t-test. **P<0.01.

B) qPCR of day-3 artery ECs derived from control and single *SOX7*-, *SOX17*-, or *SOX18*-knockdown H1 CRISPRi hPSC lines. Gene expression normalized to levels in control hPSC-derived artery ECs, which was set as “1.0”. Statistics: unpaired t-test. Error bars: SEM. n.s.: not significant. *P<0.05. **P<0.01.

C) qPCR of day-3 artery ECs generated from H1 control vs. *SOXF* TKD hPSCs. Gene expression normalized to control artery ECs. Percentages indicate remaining gene expression in *SOXF* TKD artery ECs, relative to controls. Statistics: unpaired t-test. Error bars: standard error of the mean (SEM).

1213 D) *Left*: scRNA-seq of day-3 artery ECs, day-3 primed ECs and day-4 vein ECs generated
1214 from either H1 control or *SOXF* TKD CRISPRi hPSCs. *Right*: *PECAM1*+ ECs and
1215 *PECAM1*- mesenchymal cells were detected. The same scRNAseq datasets are shown
1216 here as in **Fig. 4C**, except that mesenchymal cells were computationally excluded in
1217 **Fig. 4C**.
1218 E) qPCR of day-3 artery ECs and day-4 vein ECs generated from H1 control vs. *SOXF*
1219 TKD CRISPRi hPSCs. Gene expression normalized to reference
1220 gene *YWHAZ* (i.e., *YWHAZ* = 100%). Statistics: unpaired t-test. Error bars: SEM.
1221 *P<0.05. **P<0.01.
1222 F) Differentially expressed genes between day-3 primed ECs generated from H1 control
1223 vs. *SOXF* TKD CRISPRi hPSCs are colored.
1224 G) Gene Set Enrichment analysis (GSEA)[97] of scRNA-seq data from H1 CRISPRi control
1225 vs. *SOXF* TKD hPSCs differentiated into day-3 artery ECs, day-3 primed ECs, and day-
1226 4 vein ECs. Color represents the $\log_{10}$-transformed adjusted P-value.
1227 H) *Left*: OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 primed
1228 ECs and day-4 vein ECs. *Right*: scRNA-seq of day-3 artery ECs, day-3 primed ECs, and
1229 day-4 vein ECs generated from H1 control vs. *SOXF* TKD hPSCs. *SOXF* genes are
1230 required for *CDH5* (*VE-CADHERIN*) expression in primed ECs.
1231
1232 **Figure S10: Chromatin hallmarks of human arteriovenous identity**
1233 A) Image of a 2 day-post-fertilization (dpf) zebrafish bearing a *Sema3g* -13 kB enhancer
1234 driving *GFP* expression.
1235 B) OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 artery ECs
1236 and day-4 vein ECs.
1237 C) OmniATAC-seq, CUT&RUN, and bulk RNA-seq of H1 hPSC-derived day-3 artery ECs
1238 and day-4 vein ECs.

1239 **STAR METHODS**

1240

1241 **Key Resources Table**

Reagent or Resource	Source	Identifier
Antibodies		
CD144 FITC antibody (used 1:50 for *in vitro* studies)	BD Biosciences	560411
CD144 Alexa Fluor 647 antibody (used 1:50 for *in vitro* studies)	BD Biosciences	561567
DLL4 APC antibody (used 1:5 for *in vitro* studies)	BioLegend	346508
CXCR4 PE-Cy7 antibody (used 1:50 for *in vitro* studies)	BD Biosciences	560669
CD73 APC antibody (used 1:10 for *in vitro* studies)	BD Biosciences	560847
CD73 PE antibody (used 1:10 for *in vitro* studies)	Biolegend	344004
BST2/CD317 PE-Cy7 antibody (used 1:20 for *in vitro* studies)	BioLegend	348416
SOX17 antibody (used 1:1000 for *in vitro* studies and 1:200 for *in vivo* studies)	R&D Systems	AF-1924
ERG antibody (used 1:500 for *in vivo* studies)	Abcam	ab92513
VEGFR2 antibody (used 1:125 for *in vivo* studies)	R&D Systems	AF644
SMAα antibody Cy3 (used 1:500 for *in vivo* studies)	Sigma	C6198
GFP antibody (used 1:300 for *in vivo* studies)	Abcam	ab13970
GFP antibody (used 1:1000 for *in vivo* studies)	Aves Labs	NC9510598
Goat anti-mouse IgG secondary antibody Alexa Fluor 488 (used 1:500-1:1000 for *in vitro* studies)	Jackson ImmunoResearch	115-545-003
Donkey anti-chicken IgY secondary antibody Alexa Fluor 488 (used 1:250-1:400 for *in vivo* studies)	Thermo Fisher Scientific	A78948

Donkey anti-goat IgG secondary antibody Alexa Fluor 594 (used 1:500 for *in vivo* studies)	Thermo Fisher Scientific	A32758
Donkey anti-rabbit IgG secondary antibody Alexa Fluor 647 (used 1:500 for *in vivo* studies)	Thermo Fisher Scientific	A32795
Bacterial and viral strains		
Mix & Go! *E. coli* Competent Cells	Zymo Research	T3020
Chemicals, peptides and recombinant proteins		
mTesR Plus medium	STEMCELL Technologies	100-0276
Essential 8 medium	Thermo Fisher Scientific	A1517001
StemFlex medium	Thermo Fisher Scientific	A3349401
Penicillin/streptomycin	Thermo Fisher Scientific	15140163
Geltrex LDEV-free, hESC-qualified, reduced growth factor basement membrane matrix	Thermo Fisher Scientific	A1413302
Versene (EDTA-based dissociation buffer)	Thermo Fisher Scientific	15040066
Accutase (dissociation buffer)	Thermo Fisher Scientific	00-4555-56
TrypLE Express (dissociation buffer)	Thermo Fisher Scientific	12604013
EDTA (for flow cytometry buffer)	Thermo Fisher Scientific	15575020
DMEM/F12 + GlutaMAX	Thermo Fisher Scientific	10565042
IMDM + GlutaMAX	Thermo Fisher Scientific	31980-097
F12 + GlutaMAX	Thermo Fisher Scientific	31765-092
Polyvinyl alcohol	Sigma	P8136-250G
Chemically defined lipid concentrate	Thermo Fisher Scientific	11905-031
1-thioglycerol	Sigma	M6145-100ML
Recombinant human insulin	Sigma	11376497001
Human transferrin	Sigma	10652202001
Recombinant human Activin A	R&D Systems	338-AC-050
Recombinant human BMP4	R&D Systems	314-BP-050
Recombinant human FGF2	R&D Systems	233-FB-01M
Recombinant human VEGF	R&D Systems	293-VE-0500
CHIR99201	Tocris	4423

GDC-0941	Cellagen Technology	C4321-25
XAV939	Tocris	3748
SB505124	Tocris	3263
Forskolin	Tocris	1099
Ascorbic acid-2-phosphate (AA2P, Vitamin C analog)	Sigma	49752-10G
RO4929097	Cellagen Technology	C7649-10
DMH1	Tocris	4126
PD0325901	Tocris	4192
Thiazovivin	Tocris	3845
Dimethyl sulfoxide (DMSO)	Sigma	D2650
UltraPure DNase/RNase-free distilled water (H_2O)	Thermo Fisher Scientific	10977023
Bovine albumin fraction V (7.5% solution), to reconstitute recombinant growth factors	Thermo Fisher Scientific	15260037
Endothelial Cell Growth Medium 2 (EGM2)	Lonza	CC-3162
Triton X-100	Sigma	X100-500ML
PBS, without Ca^{2+} or Mg^{2+}, for non-virological experiments	Thermo Fisher Scientific	10010049
DPBS, without Ca^{2+} or Mg^{2+}, for virological experiments	Thermo Fisher Scientific	14190144
DAPI, for *in vitro* studies	Thermo Fisher Scientific	D1306
DAPI, for *in vivo* studies	Sigma	D9542-1MG
DMEM, for lentiviral packaging	Thermo Fisher Scientific	10566024
Fetal bovine serum, for lentiviral packaging	Atlanta Biologicals (later sold by R&D Systems)	S11550H
FuGENE HD Transfection Reagent, for lentiviral packaging	Promega	E2311
Poly-L-lysine, for lentiviral packaging	R&D Systems	3438-200-01
Opti-MEM, for lentiviral packaging	Thermo Fisher Scientific	31985070
Wizard Plus SV Minipreps DNA Purification System	Promega	A1460
Quick Ligation Kit	New England Biolabs	M2200
Hygromycin B	Thermo Fisher Scientific	10687010
Puromycin	Thermo Fisher Scientific	A1113803

Bovine serum albumin (BSA)	Sigma	A2153
Fluoromount-G mounting medium, for mouse embryology	SouthernBiotech	0100-01
32% paraformaldehyde (PFA), for mouse embryo immunostaining	Electron Microscopy Sciences	15714
32% paraformaldehyde (PFA), for mouse embryo *in situ* hybridization	Fisher Scientific	50-980-495
Methanol, for mouse embryo *in situ* hybridization	Fisher Scientific	A412-1
Tween 20, for mouse embryo *in situ* hybridization	Sigma	P9416-100ML
Glycerol, for mouse embryology	Thermo Fisher Scientific	327255000
Sucrose, for mouse embryology	Sigma	S0389
Normal donkey serum, freeze-dried, for mouse embryology	Jackson ImmunoResearch	102644-006
Tissue-Plus optimal cutting temperature (OCT) embedding medium, for mouse embryology	Fisher Scientific	23-730-571
(*Z*)-4-hydroxytamoxifen (4OHT), for *Aplnr-CreER* lineage tracing	Cayman Chemical	14854
(*Z*)-4-hydroxytamoxifen (4OHT), for *Sox17-Cre;Aplnr-DreER* lineage tracing	Sigma	H7904
Ethanol, for lineage tracing	Fisher Scientific	BP2818500
Corn oil, for lineage tracing	Sigma	C8267
Critical commercial assays		
Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3	10x Genomics	PN-1000075
Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1	10x Genomics	PN-1000121
LEGENDScreen PE-Conjugated Human Antibody Plates	BioLegend	700007
RNeasy Plus Mini Kit	Qiagen	10010049

High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	4368814
SensiFAST SYBR Lo-ROX Kit	Thomas Scientific	BIO-94050
Quick-RNA Microprep Kit	Zymo Research	R1051
Tagment DNA Enzyme 1 (TDE1)	Illumina	15027865
Tagment DNA Buffer	Illumina	15027866
MinElute PCR Purification Kit, for Omni-ATAC-seq	Qiagen	28204
AMPure XP Beads, for Omni-ATAC-seq	Beckman Coulter	A63880
CellBIND polystyrene CellSTACK chamber	Corning	3311
Deposited data		
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 1 mid primitive streak, day 2 lateral mesoderm, day 3 artery ECs (unsorted cell population), day 3 ($CD144^{+}$ $DLL4^{+}$ $CD73^{lo/-}$ FACS-purified), day 3 primed ECs (unsorted cell population), day 3 primed ECs ($CD144^{+}$ FACS-purified), and day 4 vein ECs ($CD144^{+}$ $DLL4^{-}$ $CD73^{hi}$ FACS-purified)	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	This study
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 4 vein ECs (unsorted cell population) using the Ang et al. protocol	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	NCBI PRJNA837932 (Ang et al.[48])
Single-cell RNA-seq of CRISPRi control vs. *SOX7/SOX17/SOX18*-knockdown hPSCs differentiated into day 3 artery ECs, day 3 primed ECs, and day 4 vein ECs	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	This study

Single-cell RNA-sequencing dataset of hPSCs differentiated into day 6 artery ECs using the Pan et al. protocol	Wei Kong's, Xi Wang's, and Kai Wang's laboratories (Peking University)	NCBI PRJNA1114402 (Pan et al.[58])
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 6 vein ECs using the Pan et al. protocol	Wei Kong's, Xi Wang's, and Kai Wang's laboratories (Peking University)	NCBI PRJNA1114402 (Pan et al.[58])
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 6 ECs using the Zhang et al. protocol	Andrew Baker's laboratory (University of Edinburgh)	NCBI GSE131736 (McCracken et al.[62])
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 7 ECs using the McCracken et al. protocol	Andrew Baker's laboratory (University of Edinburgh)	NCBI GSE131736 (McCracken et al.[62])
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 12 ECs using the Paik et al. protocol	Joseph Wu's laboratory (Stanford University)	NCBI GSE116555 (Paik et al.[60])
Single-cell RNA-sequencing dataset of hPSCs differentiated into day 14 ECs using the Nikolova et al. protocol	Josef Penninger's, J. Gray Camp's, and Barbara Treutlein's laboratories (ETH Zurich, IMBA, and Roche)	Array Express E-MTAB-14807 (Nikolova et al.[59])
Single-cell RNA-sequencing dataset of human fetal vasculature	Vincenzo Calvanese's and Hanna Mikkola's laboratories (University of California, Los Angeles)	NCBI GSE162950 (Calvanese et al.[94])
Bulk-population RNA-sequencing dataset of hPSCs differentiated into day 1 mid primitive streak, day 2 lateral mesoderm, and day 3 primed ECs ($CD144^+$ FACS-purified)	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	This study
Bulk-population RNA-sequencing dataset of hPSCs differentiated into day 3 artery ECs ($CD144^+$ $DLL4^+$ $CD73^{lo/-}$	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	NCBI PRJNA837932 (Ang et al.[48])

FACS-purified) and day 4 vein ECs (CD144^{+} DLL4^{-} CD73^{hi} FACS-purified)		
OmniATAC-seq dataset of hPSCs differentiated into day 3 artery ECs (CD144^{+} DLL4^{+} CD73$^{lo/-}$ FACS-purified), day 3 primed ECs (CD144^{+} FACS-purified), and day 4 vein ECs (CD144^{+} DLL4^{-} CD73^{hi} FACS-purified)	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	This study
CUT&RUN dataset of hPSCs differentiated into day 3 artery ECs (CD144^{+} DLL4^{+} CD73$^{lo/-}$ FACS-purified), day 3 primed ECs (CD144^{+} FACS-purified), and day 4 vein ECs (CD144^{+} DLL4^{-} CD73^{hi} FACS-purified)	Kyle Loh's and Lay Teng Ang's laboratories (Stanford University)	This study
Experimental models: cell lines		
H1 hESCs	WiCell	WiCell, WA01 (Thomson et al.[98])
H1 *NR2F2-2A-GFP; SOX17-2A-mPlum* knock-in reporter hESCs	This study	N/A
H1 CRISPRi-expressing hESCs	Richard She's and Jonathan Weissman's laboratories	N/A
WTC11 hiPSCs	Coriell Institute for Medical Research	Coriell Institute for Medical Research, GM25256, Kreitzer et al.[99]
SUN004.2 *CAG-mScarlet* hiPSCs	Hiro Nakauchi's laboratory	Ang et al.[48]
HEK 293T/17 cells	ATCC	CRL-11268
Experimental models: organisms/strains		
Wild-type CD-1 mice	Charles River	022
Sox17-2A-iCre mouse (otherwise known as *Sox17-Cre*)	Heiko Lickert's laboratory	Engert et al.[66]
Aplnr-CreER mouse (otherwise known as *Apj-CreER*)	Kristy Red-Horse's laboratory	Chen et al.[20]

Aplnr-DreER mouse (otherwise known as *Apj-DreER*)	Sophie Astrof's laboratory	Ramirez et al.[70]
ROSA26-LSL-zsGreen mouse (otherwise known as Ai6)	The Jackson Laboratory (originally developed by the Allen Brain Institute)	JAX stock 007906 (Madisen et al.[95])
ROSA26-LSL-tdTomato mouse (otherwise known as Ai14)	The Jackson Laboratory (originally developed by the Allen Brain Institute)	JAX stock 007914 (Madisen et al.[95])
RC::RLTG mouse	The Jackson Laboratory (originally developed by Patricia Jensen's laboratory)	JAX stock 026931 (Plummer et al.[71])
mTmG mouse	The Jackson Laboratory (originally developed by Liqun Luo's laboratory)	JAX stock 007676 (Muzumdar et al.[96])
Sema3g -13kB enhancer-LacZ mouse	Len Pennacchio's and Axel Visel's laboratories	Kosicki et al.[82]
Dll4 -12 kB enhancer-LacZ mouse	Sarah De Val's laboratory	Sacilotto et al.[78]
Wild-type AB strain zebrafish	Zebrafish International Resource Center	ZL1
Cxcr4 +135 enhancer-GFP; *kdrl-HRAS-mCherry* zebrafish	Sarah De Val's laboratory	Nornes et al.[79]
Oligonucleotides		
Primers	See **Method Details** section	This study
Pool of HCR3 *in situ* hybridization probes for human *APLNR* mRNA	Molecular Instruments	Custom order for human *APLNR* mRNA (NM_005161.6)
Pool of HCR3 *in situ* hybridization probes for zebrafish *sox7* mRNA	Molecular Instruments	Custom order
Pool of HCR3 *in situ* hybridization probes for zebrafish *aplnrb* mRNA	Molecular Instruments	Custom order
Recombinant DNA		
mU6-sgRNA 1; hU6-sgRNA 2; EF1A-PuroR-2A-GFP backbone plasmid, fragment 1	Addgene (originally developed by Jonathan Weissman's laboratory)	Addgene, 187243 (Guna et al.[100])
mU6-sgRNA 1; hU6-sgRNA 2; EF1A-PuroR-2A-GFP backbone plasmid, fragment 2	Addgene (originally developed by Jonathan Weissman's laboratory)	Addgene, 187239 (Guna et al.[100])

mU6-sgRNA; EF1A-HygroR plasmid	Karina Smolyar and Jonathan Weissman	This study
pVSV-G plasmid	Owen Witte's laboratory	Goldstein et al.[101]
pMDL plasmid	Owen Witte's laboratory	Goldstein et al.[101]
pREV plasmid	Owen Witte's laboratory	Goldstein et al.[101]
Software and algorithms		
ImageJ/Fiji	[102]	https://imagej.net/software/fiji/
Seurat v3	[93]	https://satijalab.org/seurat/
Seurat v4	[103]	https://satijalab.org/seurat/
RStudio	RStudio Team	https://www.rstudio.com/
Cell Ranger	10x Genomics (ref. [104])	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
EnsDb.Hsapiens.v86	[105]	https://bioconductor.org/packages/release/data/annotation/html/EnsDb.Hsapiens.v86.html
DESeq2	[106]	https://github.com/mikelove/DESeq2
clusterProfiler	[107]	https://guangchuangyu.github.io/software/clusterProfiler/
ShinyCell	[108]	https://github.com/SGDDNB/ShinyCell
FastQC	[109]	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Trim Galore	[110]	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Kallisto	[111]	https://pachterlab.github.io/kallisto/
EdgeR	[112]	http://bioconductor.org/packages/release/bioc/html/edgeR.html
Voom	[113]	https://rdrr.io/bioc/limma/man/voom.html
Limma	[114]	https://bioconductor.org/packages/release/bioc/html/limma.html
Dplyr	[115]	https://dplyr.tidyverse.org/
ggplot2	[116]	https://ggplot2.tidyverse.org/
Tidyverse	[117]	https://www.tidyverse.org/

FastQ Screen	118	https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/
NGmerge	119	https://github.com/harvardinformatics/NGmerge
SAMBLASTER	120	https://github.com/GregoryFaust/samblaster
SAMtools	121	https://github.com/samtools/samtools
deepTools	122	https://github.com/deeptools/deepTools
tximport	123	https://github.com/thelovelab/tximport
Diffbind	124	https://bioconductor.org/packages/release/bioc/html/DiffBind.html
HOMER	125	http://homer.ucsd.edu/homer/

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Lay Teng Ang (layteng@stanford.edu).

Materials availability

H1 *SOX17-2A-mPlum*; *NR2F2-2A-GFP* double reporter hESCs were generated as part of this study, and will be made freely available upon request and the completion of applicable Material Transfer Agreements.

Data and code availability

Bulk-population RNA-seq, single-cell RNA-seq, bulk-population ATAC-seq, and bulk-population CUT&RUN datasets generated in this study will be made available upon acceptance. An interactive portal to access the single-cell RNA-seq dataset of hPSC differentiation into artery and vein ECs is available at https://anglab.shinyapps.io/artery-vein-scrna-seq/. Computational scripts used for genomics analyses conducted in this study are available at GitHub: https://github.com/LayTengAngLab/Prevein_endothelial.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

All cells in this study were cultured in standard incubator conditions (20% O_2, 5% CO_2 and 37 °C).

Human pluripotent stem cell lines

Wild-type H1 hESCs[98] (obtained from WiCell; XY biological sex), H1 *NR2F2-2A-GFP; SOX17-2A-mPlum* knock-in reporter hESCs (generated in this study; XY biological sex), CRISPRi-expressing H1 hESCs (generated in this study; XY biological sex), wild-type WTC11 hiPSCs[99] (obtained from the Coriell Institute for Medical Research; XY biological sex), and SUN004.2 *CAG-mScarlet* hiPSCs[48] (obtained from Hiro Nakauchi's laboratory; XY biological sex) were used in this study.

All hPSC lines in this study, except for WTC11 and H1 CRISPRi hPSCs, were cultured in mTeSR Plus medium (STEMCELL Technologies). WTC11 hPSCs were cultured in Essential 8 medium[126] (Thermo Fisher Scientific). H1 CRISPRi hPSCs were cultured in StemFlex medium (Thermo Fisher Scientific).

Methods to culture undifferentiated hPSCs have been described previously[127]. In brief: undifferentiated hPSCs were propagated in mTeSR Plus medium (STEMCELL Technologies) + 1% penicillin/streptomycin in monolayer cultures, on Geltrex basement membrane matrix-coated plates (described below). To maintain undifferentiated hPSCs, mTeSR Plus was changed either every day or every other day as per the manufacturer's recommendations. In order to maintain cultures of undifferentiated hPSCs, when they became partially confluent, undifferentiated hPSCs were passaged by treating them for 7 minutes with Versene (Thermo Fisher; an EDTA-based dissociation buffer) at room temperature. Subsequently, Versene was removed, mTeSR Plus was added, and then hPSCs were manually scraped off the plate to generate clumps. hPSC clumps were then seeded onto new plates that had been precoated with Geltrex basement membrane matrix (described below) in mTeSR Plus medium + 1% penicillin/streptomycin. To reiterate, during Versene-based maintenance passaging of undifferentiated hPSCs as clumps, ROCK inhibitor was not added.

Other cell lines

HEK 293T/17 cells (obtained from ATCC; isolated from tissue of XX biological sex) and Vero C1008 cells (obtained from ECACC; isolated from tissue of XX biological sex) were used in this study.

Mouse models

Wild-type CD-1 mice (Charles River, catalog number 022), *Sox17-Cre* mice (provided by Heiko Lickert's laboratory[66]), *Aplnr-CreER* mice (provided by Kristy Red-Horse's laboratory[20]), *Aplnr-DreER* mice (provided by Sophie Astrof's laboratory[70]), *ROSA26-LSL-zsGreen* mice (Ai6; JAX stock 007906, and developed by the Allen Brain Institute[95]), *ROSA26-LSL-tdTomato* mice (Ai14; JAX stock 007914, and developed by the Allen Brain Institute[95]), *RC::RLTG* mice (JAX stock 026931, and developed by Patricia Jensen's laboratory[71]), *Sema3g -13kB enhancer-LacZ* mice (developed by Len Pennacchio's and Axel Visel's laboratories[82]), and *Dll4 -12 kB enhancer-LacZ* mice (developed by Sarah De Val's laboratory[78]), were used in this study.

Zebrafish models

Wild-type AB zebrafish and *Cxcr4 +135 enhancer-GFP*; *kdrl-HRAS-mCherry* zebrafish (developed by Sarah De Val's laboratory[79]) were used in this study.

METHOD DETAILS

1315 **Data processing and visualization**
1316 Standard plots were prepared with Microsoft Excel, Microsoft PowerPoint, or GraphPad Prism.
1317 Flow cytometry data were visualized with FlowJo. Microscope images were visualized with
1318 Fiji[102]. Genomics data were respectively wrangled and plotted using dplyr[115] and ggplot2[116] in
1319 the tidyverse[117], in the RStudio environment. Genomics tracks were visualized with the
1320 Integrated Genomics Viewer (IGV)[128], and evolutionary conservation was depicted using the
1321 Phastcons 20-way mammalian conservation track. Color palettes were chosen with the
1322 assistance of https://colorbrewer2.org/.
1323
1324 **Mouse husbandry**
1325 Adult mice of the indicated genotypes were mated to generate timed pregnancies. Females
1326 were checked each morning for a vaginal plug; noon on the day a plug was observed was
1327 defined as embryonic day 0.5 (E0.5).
1328
1329 **Lineage tracing**
1330 • ***Sox17-Cre;Aplnr-DreER* lineage tracing**: As described previously[129], 25 mg of (*Z*)-4-
1331 hydroxytamoxifen (4OHT; Sigma, H7904) was dissolved in 1250 µL ethanol (Fisher
1332 Scientific, BP2818500) by vortexing and heating at 60 °C. This yielded a 20 mg/mL stock of
1333 4OHT, which was aliquoted and stored at -20 °C. Prior to dosing mice, 50 µL aliquots
1334 (containing 1 mg of 4OHT) were heated for 10 minutes at 65 °C, and then combined with
1335 pre-warmed corn oil (250 µL, Sigma, C8267). This mixture of 4OHT and corn oil was
1336 thoroughly vortexed. Pregnant females were intraperitoneally injected at the specified
1337 labeling timepoint with 1 mg of 4OHT per mouse.
1338 • ***Aplnr-CreER* lineage tracing**: 25 mg of (*Z*)-4-hydroxytamoxifen (4OHT; Cayman Chemical,
1339 14854) was dissolved at 10mg/mL in ethanol (Fisher Scientific, BP2818500) by vortexing,
1340 and then was aliquoted and stored at -80 °C. Pregnant female mice were intraperitoneally
1341 injected with 28 mg 4OHT per kg of mouse weight at the specified labeling timepoint.

1342 ***In situ* hybridization and immunostaining of whole-mount mouse embryos**
1343 Fluorescent *in situ* hybridization (FISH) of whole-mount mouse embryos was performed using
1344 hybridization chain reaction v3.0 (HCR3)[130], and some instances, immunostaining was
1345 simultaneously performed. HCR3 was performed as per the Molecular Instruments protocol
1346 (https://www.molecularinstruments.com/hcr-v3-protocols), and is briefly summarized here.
1347
1348 Mouse embryos were dissected in ice-cold 4% paraformaldehyde (Fisher Scientific, 50-980-
1349 495, diluted in PBS). They were then subsequently fixed overnight in 4% paraformaldehyde,
1350 sequentially dehydrated in methanol solutions of increasing concentration (Fisher Scientific,
1351 A412-1), and then incubated in 100% methanol overnight.
1352
1353 Embryos were subsequently permeabilized in PBS + 0.1% Triton X-100 for 1 hour at room
1354 temperature, and then blocked in blocking buffer (PBS + 0.05% Triton X-100 + 5% BSA)
1355 overnight at 4 °C. Embryos were then stained with primary antibodies (diluted in blocking buffer)
1356 for 24-72 hours at 4 °C, washed three times with PBS, and stained with secondary antibodies
1357 (diluted in blocking buffer) for 24 hours at 4 °C. Subsequently, embryos were washed three
1358 times with PBS, and then HCR3 was performed as per the Molecular Instruments protocol.
1359 Hybridization mRNA probes, amplifiers, and buffers were obtained from Molecular Instruments.

Embryos were incubated in DAPI + SSCT (sodium chloride, sodium citrate, and Tween buffer) prior to mounting. Images were captured on an Olympus FV3000 confocal microscope.

Immunostaining of whole-mount mouse embryos

Mouse embryos were dissected in ice-cold PBS, and then fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15714; diluted in PBS) for 1 hour at 4 °C. Subsequently, embryos were washed in PBS and stained with primary antibodies (diluted in PBS + 0.1% Triton X-100) overnight at 4 °C. Following three washes in PBS + 0.1% Triton X-100, embryos were incubated in secondary antibodies (diluted in PBS + 0.1% Triton X-100) overnight at 4 °C. Embryos were then washed in PBS + 0.1% Triton X-100, counterstained with 1 µg/ml DAPI (Sigma, D9542), and washed twice in PBS. Embryos were then equilibrated sequentially in 25%, 50%, and 75% glycerol in PBS for 30 minutes at room temperature without agitation. Embryos were then mounted in Fluoromount-G (SouthernBiotech, 0100-010) and imaged on a Zeiss LSM 980 confocal microscope.

Immunostaining of cryosectioned mouse embryos

Mouse embryos were dissected in ice-cold PBS, and then fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15714; diluted in PBS) for 1 hour at 4 °C. Subsequently, embryos were washed in PBS, and dehydrated in 30% sucrose at 4°C overnight. Embryos were embedded in OCT medium (Fisher Scientific, 23-730-571), and then frozen overnight at -80 °C. Embryos were sectioned to a thickness of 20 µm and incubated at room temperature for 30 minutes. Sections were subsequently permeabilized in PBS + 0.1% Triton X-100 for 15 minutes at room temperature and blocked in 5% donkey serum + PBS + 0.05% Triton X-100 at room temperature. Sections were then stained with primary antibodies (diluted in blocking buffer) at 4 °C overnight, washed with PBS + 0.05% Triton X-100. They were then stained with secondary antibodies (diluted in blocking buffer) at room temperature, followed by additional washes in PBS + 0.05% Triton X-100, counterstained with 1 µg/ml DAPI (Sigma, D9542) for 10 minutes at room temperature, and then washed with PBS. Slides were then mounted with Fluoromount-G (SouthernBiotech, 0100-010) and imaged on a Zeiss LSM 980 confocal microscope.

Quantification of mouse embryo images

Fiji[102] was used to generate maximum-intensity projections from z-stacks of whole-mount mouse embryos or sectioned mouse embryos.

- Sox17 immunostaining: Erg+ or Sox17+ cells were manually counted in the dorsal aorta (DA) and cardinal vein (CV). The percentage of Erg+ cells that expressed Sox17 is shown.
- *Sox17-Cre;Aplnr-DreER* lineage tracing: Erg+ or GFP+ cells were manually counted in the DA and CV. The percentage of Erg+ cells that expressed GFP is shown.
- *Aplnr-CreER* lineage tracing: Erg+, Sox17+, or tdTomato+ cells were manually counted in the DA and CV. The percentage of Erg+ cells that co-expressed both Sox17 and tdTomato is shown.

Functional testing of *Dll4* enhancer element in mouse embryos

A stable transgenic mouse line bearing a genomic integration of the *Dll4* -12 kb enhancer element driving *LacZ* reporter expression, known as *Tg(Dll4-12:lacZ)*, was generated as previously described[78]. Embryos were fixed in 2% paraformaldehyde + 0.2% glutaraldehyde +

1405 1x PBS for 60 minutes. After fixation, embryos were rinsed in 0.1% sodium deoxycholate, 0.2%
1406 Nonidet P-40, 2 mM $MgCl_2$ and 1x PBS, then stained for 2-24 hours in 1 mg/ml 5-bromo-4-
1407 chloro-3-indolyo-β-D-galactoside solution (X-gal) containing 5 mM potassium ferrocyanide, 5
1408 mM ferricyanide, 0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2 mM $MgCl_2$, and 1x PBS.
1409 After staining, embryos were rinsed through a series of 1x PBS washes, then fixed overnight
1410 in 4% paraformaldehyde at 4°C. All embryos were imaged using a Leica M165C stereo
1411 microscope equipped with a ProGres CF Scan camera and CapturePro software (Jenoptik).
1412
1413 **Functional testing of *Sema3g* enhancer element in mouse embryos**
1414 The activity of the *Sema3g* -13 kb enhancer element was tested in mouse embryos through a
1415 transgenic reporter assay by Len Pennacchio's and Axel Visel's laboratories, and was reported
1416 as part of the VISTA Enhancer Browser[82,83] (VISTA Enhancer Browser ID hs2179:
1417 https://enhancer.lbl.gov/vista/element?vistaId=hs2179&alleleId=0&backbone=hZR&stage=e1
1418 1.5). In brief, a reporter construct containing the *Sema3g* -13 kb enhancer element and the
1419 *Hsp68* minimal promoter driving *LacZ* reporter expression was randomly integrated into the
1420 mouse genome; whole-mount staining of mouse embryos for *LacZ* reporter expression was
1421 then performed[82,83].
1422
1423 **Fluorescent *in situ* hybridization of zebrafish embryos**
1424 FISH of whole-mount zebrafish embryos was performed using HCR3[130].
1425
1426 Wild-type AB strain embryos were fixed overnight with 4% paraformaldehyde (BT-Fix) in PBS
1427 at 4 °C, dehydrated through a sequential ethanol series, and stored at -20°C. Embryos were
1428 rehydrated and washed three times with PBT (PBS + 0.2% bovine serum albumin + 0.2%
1429 Tween 20). A prehybridization step was done using hybridization buffer (30% formamide + 5x
1430 SSC + 9 mM citric acid at pH 6.0 + 0.1% Tween 20 + 50 µg/ml heparin + 1x Denhardt's solution,
1431 + 10% dextran sulfate) for 30 mins at 37 °C.
1432
1433 Following this, each 2 pmol of each HCR3 probe was combined with 500 µL of hybridization
1434 buffer and zebrafish embryos were incubated overnight in this solution at 37°C, while being
1435 gyrated. Following incubation, embryos were washed four times with 30% formamide + 5x SSC
1436 + 9 mM citric acid at pH 6.0 + 0.1% Tween 20 + 50 µg/mL heparin at 37°C, for 15 minutes
1437 each. This was then followed with three washes with 5x SSCT (5x SSC + 0.1% Tween 20) at
1438 room temperature for 5 minutes each. Subsequently, 30 minutes of incubation in amplification
1439 buffer (5× SSC + 0.1% Tween 20 + 10% dextran sulfate) at room temperature was performed.
1440 Hairpin probes (30 pmol each), fluorescently labeled through snap cooling of 3 µM stock
1441 solution, were added to the embryos in amplification buffer and incubated overnight at room
1442 temperature in the dark. Subsequently, samples were washed five times with 5x SSCT.
1443 Embryos were mounted in 0.6% low-melting agarose and imaged using a Nikon A1R confocal
1444 microscope and a 20× objective. Images were processed using the denoiseAI function in the
1445 NIS Elements software (Nikon) to reduce noise. Maximum intensity images were obtained with
1446 Fiji[130].
1447
1448 **Differentiation of hPSCs into artery and vein ECs**
1449 Two differentiation protocols were used throughout this study, which we refer to "Version 1"
1450 (V1, which was used for most experiments) and "Version 2" (V2, which was only used in

1451 **Extended Fig. 7A-F**). The V1 differentiation protocol has been described previously[48,127]. All
1452 RNA-seq, ATAC-seq, CUT&RUN, and LEGENDScreen profiling in this study was conducted
1453 on cells generated by the V1 differentiation protocol.
1454
1455 The V2 differentiation protocol was developed as part of this study and is more efficient at
1456 generating both artery and vein ECs than the V1 protocol. The V2 differentiation protocol is
1457 identical to the V1 protocol, with the exception that the V1 protocol entails 24 hours of lateral
1458 mesoderm differentiation, whereas the V2 protocol entails 40 hours of lateral mesoderm
1459 differentiation.
1460
1461 All differentiation was conducted in defined, serum-free CDM2 basal media[131,132]. Detailed
1462 instructions to how to prepare CDM2 are available[127].
1463
1464 • **Seeding hPSCs for differentiation** (**Step 0**). In preparation for differentiation, hPSCs were
1465 dissociated into single cells using Accutase (Thermo Fisher), because sparse seeding of
1466 cells is important for differentiation. Accutase-dissociated hPSCs were plated into recipient
1467 wells in mTeSR medium supplemented with the ROCK inhibitor thiazovivin (1 μM, Tocris; to
1468 enhance hPSC survival after passaging) onto plates precoated with Geltrex basement
1469 membrane matrix. hPSCs were seeded at a density of 25,000-50,000 cells/cm^2 were seeded
1470 (i.e., 95,000-190,000 hPSCs/well of a 12-well plate)[48,127]. To clarify, long-term maintenance
1471 of undifferentiated hPSCs entailed passaging as clumps using Versene (an EDTA-based
1472 dissociation buffer; to maintain normal karyotype), but hPSCs were dissociated using
1473 Accutase to seed single cells for differentiation. 24 hours after seeding in mTeSR + 1 μM
1474 thiazovivin, during which the hPSCs re-formed small clumps, differentiation was initiated as
1475 described below.
1476 • **Step 1: Mid primitive streak induction, 24 hours**. Day 0 hPSCs were briefly washed
1477 (DMEM/F12, Thermo Fisher) to remove all traces of mTeSR + thiazovivin. Then, they were
1478 differentiated towards mid primitive streak in CDM2 media supplemented with Activin A (30
1479 ng/mL, R&D Systems), BMP4 (40 ng/mL, R&D Systems), CHIR99021 (6 μM, Tocris), FGF2
1480 (20 ng/mL, Thermo Fisher), as previously described[48,127,132]. In both the V1 and V2 protocols,
1481 mid primitive streak induction was conducted for 24 hours.
1482 • **Step 2: Lateral mesoderm induction, 24-40 hours**. Day 1 mid primitive streak cells were
1483 briefly washed (DMEM/F12) and then differentiated towards lateral mesoderm in CDM2
1484 media supplemented with BMP4 (40 ng/mL), GDC-0941 (2.5 μM, Cellagen Technology),
1485 Forskolin (10 μM, Tocris), SB-505124 (2 μM, Tocris), VEGF (100 ng/mL, R&D Systems),
1486 XAV939 (1 μM, Tocris) and ascorbic acid-2-phosphate (AA2P; 200 μg/mL, Sigma), as
1487 previously described[48,127]. In the V1 protocol, lateral mesoderm induction was performed for
1488 24 hours. In the V2 protocol, lateral mesoderm induction was performed for 40 hours.
1489 Subsequently, lateral mesoderm was subjected to either artery EC induction (Step 3A;
1490 below) or primed EC induction (Step 3A; below).
1491 • **Step 3A: Artery EC induction, 24 hours**. Day 2 lateral mesoderm cells (V1 protocol) or
1492 day 2.67 lateral mesoderm cells (V2 protocol) were briefly washed (DMEM/F12) and then
1493 differentiated towards artery ECs in CDM2 media supplemented with Activin A (15 ng/mL),
1494 DMH1 (250 nM, Tocris), GDC-0941 (2.5 μM), VEGF (100 ng/mL), XAV939 (1 μM) and AA2P
1495 (200 μg/mL), as previously described[48,127]. In both the V1 and V2 protocols, artery EC
1496 induction was performed for 24 hours.

- **Step 3B: Primed EC induction, 24 hours**. Day 2 lateral mesoderm cells (V1 protocol) or day 2.67 lateral mesoderm cells (V2 protocol) were briefly washed (DMEM/F12) and then differentiated into primed ECs in CDM2 media supplemented with SB505124 (2 μM), DMH1 (250 nM), RO4929097 (2 μM, Cellagen Technology), VEGF (100 ng/mL), XAV939 (1 μM) and AA2P (200 μg/mL), as previously described[48,127]. In both the V1 and V2 protocols, primed EC induction was performed for 24 hours.
- **Step 4B: Vein EC induction, 24 hours**. Day 3 primed ECs (V1 protocol) or day 3.67 primed ECs (V2 protocol) were briefly washed (DMEM/F12) and then differentiated into vein ECs in CDM2 media supplemented with SB505124 (2 μM), RO4929097 (2 μM), PD0325901, (500 nM, Tocris), CHIR99021 (1 μM) and AA2P (200 mg/mL), as previously described[48,127]. In both the V1 and V2 protocols, vein EC induction was performed for 24 hours.

Detailed methods to reconstitute each differentiation-inducing small molecule and recombinant growth factor and to prepare stocks of each are available[127]. In brief, (1) all recombinant growth factors were reconstituted in PBS + 0.1% bovine albumin fraction V (both from Thermo Fisher Scientific), (2) all small molecules except for AA2P were reconstituted in DMSO (Sigma), and (3) AA2P was reconstituted in H_2O (Thermo Fisher Scientific), as described previously[127].

Large-scale differentiation of hPSCs into artery and vein ECs

Large-scale differentiation of hPSCs was performed using the V2 differentiation protocol described above.

- **Artery EC differentiation**: WTC11 hPSCs were seeded at a density of 20.5K cells/cm^2 in a Geltrex-coated 5-stack CellSTACK device (Corning), which yielded 94.5 million cells per device.
- **Vein EC differentiation**: WTC11 hPSCs were seeded at a density of 25.2K cells/cm^2 in a Geltrex-coated 5-stack CellSTACK device, which yielded 239.6 million cells per device. Alternatively, WTC11 hPSCs were seeded at a density of 17.2K cells/cm^2 in 60 15-cm dishes, which yielded 654 million cells altogether.

Cryopreservation and maintenance of hPSC-derived artery and vein ECs

After hPSC differentiation into artery ECs or vein ECs as described above, they could be maintained for at least 6 additional days *in vitro* on Geltrex-coated cell culture plates. As described previously[48], hPSC-derived artery ECs were expanded in EGM2 (Endothelial Cell Growth Medium 2, Lonza CC-3162), which was refreshed every 24 hours. By contrast, hPSC-derived vein ECs were expanded in EGM2 + SB505124 (2 μM) + RO4929097 (2 μM) + Forskolin (10 μM), which was refreshed every 24 hours.

Alternatively, as previously described[48], hPSC-derived day 3 artery ECs and day 4 vein ECs were dissociated, cryopreserved in freezing media (90% PBS + 10% DMSO), and stored in liquid nitrogen. hPSC-derived artery and vein ECs were then thawed in their respective media (EGM2 for artery ECs and EGM2 + SB505124 + RO4929097 + Forskolin for vein ECs) and cultured for up to 6 days as described above, with Thiazovivin (1 μM) added for the first 24 hours post-thawing to improve cell survival.

Flow cytometry

Cultured cells were dissociated by incubation in TrypLE Express (Thermo Fisher) for 5 minutes at 37 °C. Following dissociation, the cells were diluted with 5-10 times excess volume of FACS

buffer (PBS + 1 mM EDTA [Thermo Fisher] + 2% v/v FBS [Atlanta Bio] + 1% v/v Penicillin/Streptomycin [Thermo Fisher]) and centrifuged at 500g for 5 minutes to pellet them. Each cell pellet was then resuspended in FACS buffer and incubated with fluorescently-conjugated primary antibodies for 15-30 minutes in the dark at 4°C. After staining, cells were washed twice with FACS buffer and resuspended in 100 μL FACS buffer containing DAPI (1 μg/mL) for live/dead discrimination. Flow cytometry was conducted on a Beckman Coulter CytoFlex analyzer in the Stanford Stem Cell Institute FACS Core Facility. Flow cytometry data analysis was performed using FlowJo software. Cells were gated based on forward and side scatter area, followed by height and width parameters for doublet discrimination. Subsequently, live cells that were negative for DAPI were gated for marker analyses and calculations of population frequency. Gates to determine the percentage of cells positive for a given marker were set based on negative controls, namely cells that were not stained with an antibody or cells that were lacking the genetically encoded fluorescent reporter of interest.

In this study, we defined hPSC-derived artery and vein ECs using the following cell-surface marker combinations:

- **Artery ECs**: $CD144^+$ $CXCR4^+$ $DLL4^+$ (and, in some experiments, $CD144^+$ $DLL4^+$ $CD73^{lo/-}$). The following antibody combination was used to define arterial identity: CD144 FITC (BD Biosciences, 560411 [1:50 concentration]), DLL4 APC (Biolegend, 346508 [1:5 concentration]), and CXCR4 PE-Cy7 (BD Biosciences, 560669 [1:50 concentration]).
- **Vein ECs**: $CD144^+$ $CD317^+$ $CD73^+$ (and, in some experiments, $CD144^+$ $DLL4^-$ $CD73^{hi}$). The following antibody combination was used to define venous identity: CD144 FITC (BD Biosciences, 560411 [1:50 concentration]), CD73 APC (BD Biosciences, 560847 [1:10 concentration]), and CD317 PE-Cy7 (BioLegend, 348416 [1:20 concentration]).

The CD73 protein is encoded by the *NT5E* gene. The CD317 protein is encoded by the *BST2* gene.

Quantitative PCR

Methods for RNA extraction, reverse transcription, and qPCR have been described previously[48]. In brief, undifferentiated or differentiated hPSCs were first lysed in 350 μL of RLT Plus Buffer and RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's protocol. Second, 300 ng of total RNA was reverse transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. Third, qPCR was performed in 384-well format using the SensiFAST SYBR Lo-ROX Kit (Thomas Scientific) as previously described[48,132], using gene-specific forward and reverse primers on a QuantStudio 5 qPCR machine (Thermo Fisher). Expression of all genes was normalized to the levels of the reference gene *YWHAZ*.

Human gene	Forward primer	Reverse primer
YWHAZ	GAGCTGGTTCAGAAGGCCAAAC	CCTTGCTCAGTTACAGACTTCATGCA
HAND1	GTGCGTCCTTTAATCCTCTTC	GTGAGAGCAAGCGGAAAAG
ISL1	AGATTATATCAGGTTGTACGGGATCA	ACACAGCGGAAACACTCGAT

VEGFR2/FLK1/KDR	TTTTTGCCCTTGTTCTGTCC	TCATTGTTCCCAGCATTTCA
ETV2	CCGACGGCGATACCTACTG	CGGTGGTTAGTTTTGGGGCAT
SCL/TAL1	CAAAGTTGTGCGGCGTATCTT	TCATTCTTGCTGAGCTTCTTGTC
LMO2	ATTGGGGACCGCTACTTC	GCCCAAAAAGCCTGAGATAGT
FLI1	ACCTCCCACACCGACCAAT	GGACTTTTGTTGAGGCCAGAA
CDH5/VE-CADHERIN	AACGAGCAGGGCGAGTTCACCTTC	TAGGTGACCAGCTGCTCGTGGATC
PECAM1/CD31	AACAGTGTTGACATGAAGAGCC	TGTAAAACAGCACGTCATCCTT
DLL4	GTCTCCACGCCGGTATTGG	CAGGTGAAATTGAAGGGCAGT
CXCR4	CACCGCATCTGGAGAACCA	GCCCATTTCCTCGGTGTAGTT
EFNB2	AAGGACTGGTACTATACCCACAG	TGTCTGCTTGGTCTTTATCAACC
SOX7	AGCCGGAGCAGACCTTCTT	GCCGGGGAGTAATAGGCAG
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
SOX18	AAGCGTGGAAGGAGCTGAAC	CGCGGCCGGTACTTGTAGTT
NR2F2	GCCATAGTCCTGTTCACCTCA	AATCTCGTCGGCTGGTTG
APLNR	CTCTGGACCGTGTTTCGGAG	GGTACGTGTAGGTAGCCCACA
FLRT2	CGCTGCGACAGGAACTTTG	TGGAGGTAGAGTACGGTTACG
NRP2	GCTGGCTATATCACCTCTCCC	TCTCGATTTCAAAGTGAGGGTTG
NT5E/CD73	CCAGTACCAGGGCACTATCTG	TGGCTCGATCAGTCCTTCCA

Combined immunostaining and *in situ* hybridization of cultured cells

Combined immunostaining and *in situ* hybridization of cultured cells was performed as described by Molecular Instruments (https://files.molecularinstruments.com/MI-Protocol-2%C2%BAIF-RNAFISH-GenericSolution-Rev6.pdf). First, cultured monolayer cells were fixed, permeabilized and immunostained with primary and secondary antibodies. Next, probe hybridization, amplification, and wash steps were performed using the HCR3 protocol[130]. Imaging was conducted using an FV3000 confocal microscope (Olympus).

Bulk-population RNA-seq

We performed bulk-population RNA-seq on the following cell populations that were generated from H1 hPSCs in the same biological experiment: *1*) day-0 hPSCs (entire cell population, unsorted); *2*) day-1 mid primitive streak (entire cell population, unsorted); *3*) day-2 lateral mesoderm (entire cell population, unsorted); *4*) day-3 artery ECs (FACS-purified for $CD144^+$ $DLL4^+$ $CD73^{lo/-}$ cells; 10 μM forskolin was added during day-3 artery EC induction); *5*) day-3 primed ECs (FACS-purified for $CD144^+$ cells); and *6*) day-4 vein ECs (FACS-purified for $CD144^+$ cells). Bulk-population RNA-seq datasets of day-3 artery ECs and day-4 vein ECs were previously reported by Ang et al., 2022 (NCBI PRJNA837932)[48].

Bulk-population RNA-seq was performed and computationally analyzed as described previously[48]. Cells were lysed in Zymo RNA lysis buffer, and RNA was purified using the Zymo Quick-RNA Microprep Kit (Zymo, R1051). RNA integrity was assessed by Agilent Bioanalyzer on-chip electrophoresis. High-quality RNA samples with RNA integrity number ≥ 8 underwent poly(A) enrichment, and libraries were prepared with indexed adaptors for multiplexing. RNA-

seq libraries were then sequenced on the DNBSEQ-G400 sequencer by BGI Global Genomic Services to generate 150-bp paired-end reads. To limit batch effects, all libraries were pooled prior to sequencing and distributed across multiple planes.

FastQC[109] was used to perform quality control of raw RNA-seq reads. Adapters and low-quality bases were trimmed with Trim Galore[110]; a Phred quality threshold of ≥ 33 was used, and reads shorter than 20 nucleotides after trimming were discarded. Reads from each library—which was originally sequenced on multiple lanes—were then concatenated to yield one file per library. RNA-seq reads were pseudoaligned to human reference genome hg38, and gene-level RNA-seq counts were then quantified, using Kallisto[111]. After quantification of gene-level RNA-seq counts, two separate computational workflows in the RStudio environment were used:

- **Workflow 1** (**Generation of volcano plots**): Gene counts were filtered with edgeR[112]. Counts were then transformed to $\log_2$ counts per million (CPM) using voom[113]. Differentially expressed genes were determined by limma[114], and P values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg method to control the false discovery rate.
- **Workflow 2** (**Generation of gene expression matrices**): Ensembl transcript IDs were mapped to Ensembl gene IDs using EnsDb.Hsapiens.v86[105]. Transcript-level abundance estimates were summarized to gene-level counts with tximport[123], and then imported into DESeq2[106] for differential expression analysis.

Single-cell RNA-sequencing of hPSC-derived cell populations

The 10x Genomics Chromium platform was used to perform single-cell RNA-seq profiling every 24 hours during the differentiation of H1 hPSCs into artery and vein ECs using the V1 differentiation protocol.

In one experiment, the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 was used to profile the following samples: *1*) day-0 hPSCs (entire cell population, unsorted); *2*) day-1 mid primitive streak (entire cell population, unsorted); *3*) day-2 lateral mesoderm (entire cell population, unsorted); *4*) day-3 primed EC (entire cell population, unsorted); *5*) day-3 primed EC (FACS-purified for CD144$^+$ cells); *6*) day-4 vein EC (entire cell population, unsorted); and *7*) day-4 vein EC (FACS-purified for CD144$^+$ DLL4$^-$ CD73^{hi} cells). In a second experiment, the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3.1 was used to profile the following samples: *8*) day-3 artery EC (entire cell population, unsorted); and *9*) day-3 artery EC (FACS-purified for CD144$^+$ DLL4$^+$ CD73$^{lo/-}$ cells). This scRNAseq dataset of unsorted day-4 vein ECs was previously reported by Ang et al., 2022 (NCBI PRJNA837932)[48]; all other scRNAseq datasets were generated as part of this study.

The rationale to perform scRNAseq of FACS-purified ECs was to rigorously test whether cell-surface marker combinations would enable the isolation of transcriptionally-homogeneous cell populations.

As per the manufacturer, no batch effects have been detected between the Chromium v3 and v3.1 chemistries (https://kb.10xgenomics.com/hc/en-us/articles/360047373071-Does-Cell-Ranger-distinguish-between-v3-and-v3-1-chemistry).

Sequencing libraries were prepared using Chromium Single Cell 3' GEM Gene Expression v3 or v3.1 kits as per the manufacturer's guidelines. Sequencing libraries were diluted in Buffer EB. Libraries were prepared with 10-nucleotide indices compatible with Illumina sequencers.

Computational analysis of single-cell RNA-sequencing data from hPSC-derived cell populations

scRNAseq libraries were sequenced across multiple lanes on an Illumina HiSeq 4000 sequencer by Novogene. The first and last 8 nucleotides of the i7 indices were unique and thus used for demultiplexing. FASTQ sequencing files were input into Cell Ranger[104], which was used to align reads to the hg38 reference genome (version GRCh38-2024-A), followed by filtering, barcode counting, and unique molecular identifier (UMI) counting.

Subsequent analyses were performed using Seurat v3[93] in the RStudio environment. Cell matrix files generated from Cell Ranger were imported into R using the Seurat function "Read10x_h5". For each individual scRNAseq dataset, we performed quality control by excluding dying/dead cells that *1*) exhibited low numbers of expressed genes, *2*) displayed anomalously low or high mitochondrial counts, or *3*) did not express reference genes *ACTB* or *YWHAZ*; additionally, we also *4*) computationally excluded likely doublets by removing cells that had significantly higher counts of expressed genes. High-quality single-cell transcriptomes that passed these quality control metrics were used for subsequent analyses.

Seurat objects from all scRNAseq datasets were merged using the "Merge" function in Seurat. scRNAseq data were then normalized using the "LogNormalized" function and scaled using a linear transformation in Seurat. Marker gene expression was depicted on UMAP plots[133] by coloring each single cell (i.e., a dot) according to the levels of marker gene expression. Dots were randomly ordered, without visually superimposing dots that were positive for a given marker gene, thereby avoiding visual stacking bias.

For scRNAseq analysis of control vs. *SOXF*-deficient CRISPRi hPSCs that were subject to EC differentiation, first we computationally identified ECs and excluded non-ECs from the scRNAseq dataset. This was performed via Louvain clustering, which clearly distinguished *PECAM1*+ ECs vs. *PECAM1*- non-EC clusters (**Fig. S9D**). EC clusters were used for subsequent analyses, using the aforementioned scRNAseq analysis workflows. Gene set enrichment analysis of gene ontology terms was performed on genes that were differentially expressed between control vs. *SOXF*-CRISPRi ECs using clusterProfiler[107]. Selected gene ontology terms were shown.

Data wrangling and plotting were respectively performed using dplyr[115] and ggplot2[116] in the tidyverse[117]. An interactive web browser to explore scRNAseq data was constructed using ShinyCell[108].

Comparing different endothelial differentiation protocols using single-cell RNA-seq

We analyzed scRNAseq datasets of hPSCs that were subjected to 8 different EC differentiation protocols. The 8 differentiated cell populations analyzed were:

- **Ang et al., day 3 artery ECs**: hPSCs were differentiated into artery ECs in 3 days using the V1 differentiation protocol described by Ang et al., 2022[48]. This new scRNAseq dataset was generated as part of this study.
- **Ang et al., day 4 vein ECs**: hPSCs were differentiated into vein ECs in 4 days using the V1 differentiation protocol described by Ang et al., 2022[48]. This scRNAseq dataset was deposited by the Ang et al., 2022 study[48] and is publicly available from NCBI PRJNA837932.
- **Pan et al., day 6 artery ECs**: hPSCs were differentiated into artery ECs in 6 days using the Pan et al., 2024 protocol[58]. This scRNAseq dataset was deposited by the Pan et al., 2024 study[58], and is publicly available from NCBI PRJNA1114402.
- **Pan et al., day 6 vein ECs**: hPSCs were differentiated into vein ECs in 6 days using the Pan et al., 2024 protocol[58]. This scRNAseq dataset was deposited by the Pan et al., 2024 study [58], and is publicly available from NCBI PRJNA1114402.
- **Zhang et al., day 6 ECs**: hPSCs were differentiated into ECs in 6 days using the Zhang et al., 2017 differentiation protocol[61]. This scRNAseq dataset was deposited by the McCracken et al., 2019 study[62], and is publicly available from NCBI GSE131736.
- **McCracken et al., day 7 ECs**: hPSCs were differentiated into ECs in 7 days using the McCracken et al., 2019 protocol[62]. This scRNAseq dataset was deposited by the McCracken et al., 2019 study[62], and is publicly available from NCBI GSE131736. Data from 3 experimental replicates were deposited to the Gene Expression Omnibus, and our preliminary analysis revealed batch effects among these three experimental replicates. To reduce batch effects, we selected replicate 3 for analysis, as it was the replicate that contained the largest number of cells. Of note, the original McCracken et al. study refers to cells being harvested on day 8 of differentiation[62]. However, because the first day of differentiation in their procedure entails cell seeding in hPSC medium[62], to be consistent with the nomenclature used here to describe other differentiation protocols, here we refer to these cells being differentiated for 7 days.
- **Paik et al., day 12 ECs**: hPSCs were differentiated into ECs in 12 days using the Paik et al., 2018 protocol[60]. This scRNAseq dataset was deposited by the Paik et al., 2018 study[60], and is publicly available from NCBI GSE116555.
- **Nikolova et al., day 14 ECs**: hPSCs were differentiated into vascular organoids in 14 days using the Nikolova et al., 2025 protocol[59], which in turn was modified from the Wimmer et al., 2019 protocol[134] to generate vascular organoids. This scRNAseq dataset was deposited by the Nikolova et al., 2025 study[59], and is publicly available from ArrayExpress E-MTAB-14807.

scRNAseq datasets were analyzed using Seurat v3[93]. For all scRNAseq datasets, first we selected single-cell transcriptomes that passed well-established quality control metrics, by excluding dying/dead cells that *1*) exhibited low numbers of expressed genes, *2*) displayed anomalously low or high mitochondrial counts, or *3*) did not express reference genes *ACTB* or *YWHAZ*; additionally, we also *4*) computationally excluded likely doublets by removing cells that had significantly higher counts of expressed genes. For initial analyses, we assessed all single-cell transcriptomes that passed these quality control criteria; we did not pre-select a given subpopulation of cells based on marker gene expression that might bias further analyses.

Seurat objects from all scRNAseq datasets were then merged using the “Merge” function in Seurat v3[93]. To quantify the degree of cellular heterogeneity generated by each differentiation protocol, Louvain clustering was applied at the same resolution (0.1) across all datasets, thus

decomposing each scRNAseq dataset into multiple constituent cell-types. To assign the identity of each "cell-type" within each dataset, we analyzed genes that were differentially expressed between each cell-type within a given dataset, which we annotated based on the expression of known marker genes:

- **Ang et al., day 3 artery ECs** (3 constituent clusters): artery EC (*EDN1*+ *MKI67*-), dividing artery EC (*EDN1*+ *MKI67+*), mesenchyme (*ACTC1*+ *TPM1*+)
- **Ang et al., day 4 vein ECs** (4 constituent clusters): vein EC *(FLRT2*+ *NEFH+)*, dividing vein EC (*FLRT2*+ *DGKB+),* other EC (*LNCAROD*+), mesenchyme (*ACTC1*+ *TPM1*+)
- **Pan et al., day 6 artery ECs** (4 constituent clusters): EC (*EGFL7*+ *ASPM*-), dividing EC (*EGFL7*+ *ASPM*+), mesenchyme (*IGFBP3*+ *TOP2A*-), dividing mesenchyme (*IGFBP3*+ *TOP2A*+)
- **Pan et al., day 6 vein ECs** (3 constituent clusters): EC (*KDR*+ *TOP2A*-), dividing EC (*KDR*+ *TOP2A*+), mesenchyme (*ACTC1*+)
- **Day 6 Zhang et al.** (5 constituent clusters): EC (*CLDN5*+ *UBE2C*-), dividing EC (*CLDN5*+ *UBE2C*+), mesenchyme (*MEST*+ *ACTC1*+), blood-like (*SPI1/PU.1*+), heart/kidney (*NKX2.5*+ *LHX1*+)
- **McCracken et al., day 7 ECs** (3 constituent clusters): EC (*PLVAP*+ *MKI67*-), dividing EC (*PLVAP*+ *MKI67*+), mesenchyme (*MEST*+ *TAGLN*+)
- **Paik et al., day 12 ECs** (5 constituent clusters): EC (*ECSCR*+), mesenchyme (*LUM*+ *CENPF*-), proliferating mesenchyme (*LUM*+ *CENPF*+), liver-like (*FGB+ APOA2*+ *TTR*+ *AFP*+), unknown (co-expression of epithelial marker *EPCAM* and mesenchymal marker *LUM*)
- **Nikolova et al., day 14 ECs** (3 constituent clusters): EC (*CLDN5*+), mesenchyme (*LUM*+ *TOP2A*-), dividing mesenchyme (*LUM*+ *TOP2A*+)

Assessing arteriovenous identity of cells generated from each hPSC differentiation protocol, using single-cell RNA-seq

We quantified the arteriovenous identity of hPSC-derived ECs that were profiled by aforementioned scRNAseq studies. Each dataset comprises a mixture of ECs and non-ECs at varying proportions. First, we computationally selected the EC cluster generated by each differentiation protocol, as per the above cluster annotations. (For differentiation protocols that generated "EC" and "dividing EC" clusters, all EC clusters were combined.) Then, in these EC populations obtained from distinct differentiation protocols, we analyzed the expression of arterial vs. venous marker gene modules, which are referred to as the "artery signature" or "vein signature" in this study. The Hou et al., 2022 study[24] previously demonstrated that the expression of these arteriovenous gene modules is evolutionarily conserved across ECs obtained from both human and mouse embryos:

- **Arterial gene module**: *GJA4*, *UNC5B*, *DLL4*, *MECOM*, *HEY1*, *EFNB2*, *EPAS1*, *CXCR4*, *IGFBP3*
- **Venous gene module**: *NR2F2*, *NRP2*, *APLNR*, *FLRT2*

We implemented the AddModuleScore function of Seurat v3[93] to calculate the average expression of these arterial module genes and venous module genes in hPSC-derived ECs generated from each differentiation protocol.

Single-cell RNA-seq analysis of endothelial heterogeneity in human and mouse embryos

We downloaded scRNAseq datasets of ECs isolated from the Carnegie Stage 17 human embryo (generated by Calvanese et al., 2022[94]; NCBI GSE162950, sample GSM4968834). Computational analysis was performed as described above, with the exception that Seurat v4[103] was used.

OmniATAC-seq library construction

We performed OmniATAC-seq profiling of hPSC-derived day 3 artery ECs (FACS-purified for $CD144^{+}$ $DLL4^{+}$ $CD73^{lo/-}$ cells) and day 4 vein ECs (FACS-purified for $CD144^{+}$ $DLL4^{-}$ $CD73^{hi}$ cells). A slightly modified version (https://www.med.upenn.edu/kaestnerlab/assets/user-content/documents/ATAC-seq-Protocol-(Omni)-Kaestner-Lab.pdf) of the original OmniATAC-seq protocol[49] was employed here, and is briefly summarized below.

First, we prepared resuspension buffer (10 mM Tris-HCl, pH 7.5 + 10 mM NaCl + 3 mM $MgCl_2$ + nuclease-free H_2O), cold lysis buffer (resuspension buffer + 0.1% v/v NP-40 + 0.1% v/v Tween-20 + 0.01% v/v digitonin), and wash buffer (99.9% resuspension buffer + 0.1% v/v Tween-20).

50,000 cells from each hPSC-derived cell-type were pelleted and washed with 500 µL of cold PBS, before lysis in 100 µL cold lysis buffer for 3 minutes on ice. To the cell lysate, 1 mL of cold wash buffer was added. The mixture was centrifuged at 500g for 10 minutes at 4 °C. Then, the supernatant (cytoplasm) was discarded and the pellet (nuclei) was retained. Transposition reaction mix from the Nextera DNA library prep kit (1x Tagment DNA [TD] Buffer + 1x PBS + 0.1% v/v Tween-20 + 0.01% v/v Digitonin + Tn5 Transposase [Tagment DNA Enzyme 1] + nuclease-free H_2O) was added to the pellet to resuspend nuclei. The transposition reaction was incubated at 37 °C for 30 minutes on a thermal mixer, with shaking at 1,000 rpm.

DNA was purified using the Qiagen MinElute Reaction Cleanup Kit, and then PCR-amplified using Illumina i5 and i7 index primers on a thermal cycler. Then, the libraries were purified using AMPure XP beads. All OmniATAC-seq libraries were multiplexed such that they could be sequenced as a pool on a single lane. After quality control was performed on pooled libraries, deep sequencing was performed on an Illumina HiSeq sequencer (~350 million reads/lane) and an Illumina NovaSeq S4 sequencer (~2500 million reads/lane). According to general guidelines, a minimum of 50 million reads are needed to identify accessible chromatin elements and 200 million reads are needed to identify enriched transcription factor motifs (https://www.illumina.com/techniques/popular-applications/epigenetics/atac-seq-chromatin-accessibility.html) by ATAC-seq. In this study, OmniATAC-seq libraries were sequenced at a depth of 292-1908 million raw reads per library.

OmniATAC-seq computational analysis

OmniATAC-seq data were computationally processed using the standardized ENCODE ATAC-seq analysis pipeline (https://www.encodeproject.org/atac-seq/). First, reads were aligned to human reference genome hg38 using Bowtie2[135]. Then, MACS2[136] was used to call peaks for each library. A unified peak list for each cell-type was generated by selecting only peaks that were reproducible between the two replicates. This was achieved through an irreproducible discovery rate (IDR) analysis at the threshold of 0.05 described by the ENCODE Consortium[137].

Finally, peaks that overlapped with a "black list" of artifactual regions in hg38 (https://sites.google.com/site/anshulkundaje/projects/blacklists) were removed.

Diffbind[124] was used to identify chromatin regions that exhibited >8-fold differential accessibility between artery and vein ECs. HOMER[125] was used to discover DNA motifs overrepresented in these artery- or vein-accessible elements. HOMER analyses were run on repeat-masked hg38 sequences extracted in 200-nucleotide windows centered on peak summits, using the "findMotifsGenome.pl" function.

CUT&RUN library construction

CUT&RUN profiling of H3K4me1, H3K4me3, H3K27me3, and SOX17 was conducted on H1 hPSCs differentiated into day 3 artery ECs, day 3 primed ECs, and day 4 vein ECs. For most CUT&RUN experiments, CD144+ ECs were purified by FACS to exclude any contaminating mesenchymal cells, with the exception for SOX17 CUT&RUN, as SOX17 is not expressed in mesenchymal cells. As a negative control, an isotype IgG control antibody was also separately included.

CUT&RUN profiling and library construction was performed as previously described[55]. CUT&RUN libraries were sequenced on an Illumina NovaSeq X Plus by Novogene.

CUT&RUN computational analysis

FastQC[109] and FastQ Screen[118] were used to perform quality control of raw CUT&RUN reads. Paired-end reads were merged using NGmerge[119] and aligned to human reference genome hg38 with Bowtie2[64], using the --very-sensitive mode and fragment length parameters -I 10 -X 2000. Resulting SAM files were processed with SAMBLASTER[120] to remove PCR duplicates. SAMtools[121] was then used to convert files into the sorted BAM format. Bigwig signal tracks were generated using the "bamCoverage" function from deepTools[122], with 5-nucleotide bin size and values displayed in reads per kilobase million (RPKM).

High-throughput surface marker screen by flow cytometry

The expression of 332 cell-surface markers was assessed across undifferentiated hPSCs (day 0), primitive streak (day 1), lateral mesoderm (day 2), artery ECs (day 3) and vein ECs (day 4) through the use of high-throughput flow cytometry as described previously[132]. In brief, hPSCs or their differentiated mesoderm progeny were dissociated using TrypLE Express. They were then plated into individual wells of four 96-well LEGENDScreen PE-Conjugated Human Antibody Plates (Biolegend, 700001). Each well containing a distinct antibody against a human cell-surface antigen, altogether totaling 332 unique cell-surface markers across four 96-well plates. High-throughput cell-surface marker staining was largely done as per the manufacturer's recommendations, and cells were stained with a viability dye (DAPI) prior to robotically-enabled plate-based analysis on an BD FACSCanto II (Stanford Stem Cell Institute FACS Core). Stained cells were not fixed prior to FACS analysis. LEGENDScreen data for undifferentiated H7 hPSCs (day 0) and H7-derived anterior primitive streak (day 1) were published previously[132]. LEGENDScreen data for H1-derived lateral mesoderm (day 2), H1-derived artery ECs (day 3) and H1-derived vein ECs (day 4) was generated in this study. Day 3 artery ECs and day 4 vein ECs were both co-stained with an anti-CD144 Alexa Fluor 647

antibody (BD Biosciences, 561567) to identify CD144+ ECs, and surface-marker expression was evaluated specifically in the CD144+ population.

Assembling CRISPRi constructs

sgRNAs targeting the human *SOX7*, *SOX17*, or *SOX18* genes were selected from genome-wide libraries of CRISPRi sgRNAs[138].

- *SOX7* CRISPRi sgRNA 1: TCGCCTCGCTTCGCCTGGCG
- *SOX7* CRISPRi sgRNA 2: GAAGCGAGGCGACCCGCGTG
- *SOX17* CRISPRi sgRNA: GCGACAGGCCAGAACACGGG
- *SOX18* CRISPRi sgRNA: GCGGATGGCGGTGGGGACGG

To prepare sgRNA inserts, we synthesized the following oligonucleotides in preparation for introduction into the single sgRNA plasmid (harboring BstXI and Bmtl restriction sites):

- 5' ttg + top strand sgRNA + gtttaagagc 3'
- 5' ttagctcttaaac + bottom strand sgRNA + caacaag 3'

To prepare sgRNA inserts, we synthesized the following oligonucleotides in preparation for introduction into the dual sgRNA plasmid (harboring BstXI, BsmBI, and BlpI restriction sites):

- 5' ttg + position A top strand sgRNA + tctca 3'
- 5' ctcttgaga + position A bottom strand sgRNA + caacaag 3'
- 5' gaaaggag + position B top strand sgRNA + gtttaagagc 3'
- 5' ttagctcttaaac + position B top strand sgRNA + ctcc 3'

These oligonucleotides were annealed using 20 μL of each oligonucleotide at 100 μM concentration + 10 μL of 10x annealing buffer (100 μM Tris HCl (pH 7.5), 500 mM NaCl, 10 mM EDTA (pH 8.0)) + 50 μL water. Annealing was performed at 99 °C for 5 minutes and then brought to 25 °C for 5 minutes in a thermal cycler.

To individually knock down each of these genes, we cloned each individual sgRNA into a separate *mU6-sgRNA; EF1A-HygroR* plasmid using the Quick Ligation Kit (New England Biolabs, M2200), generally using the manufacturer's protocol but with modified reaction volumes: 0.33 μL digested plasmid at 25 ng/μL concentration, 1.67 μL 2x quick ligase buffer, 0.166 μL quick ligase, and 1.33 μL annealed sgRNA at 20 nM concentration per reaction.

In experiments where we sought to simultaneously knockdown all three genes, we assembled a dual sgRNA plasmid using a previously-described strategy[100]. A dual sgRNA plasmid was assembled from two backbone fragments (Addgene, 187243 and 187239, respectively). We assembled a *mU6-SOX7-sgRNA 1; hU6-SOX17-sgRNA; EF1A-PuroR-2A*-GFP plasmid using the Quick Ligation Kit (New England Biolabs, M2200). The manufacturer's protocol was generally followed, but modified reaction volumes were used: 0.33 μL of backbone plasmid 1 at 25-33 ng/μL concentration, 0.33 μL of backbone plasmid 2 at 25-33 ng/μL concentration, 1.67 μL 2x quick ligase buffer, 0.166 μl quick ligase, 0.5 μL of annealed sgRNA 1 at 200 nM concentration per reaction, and 0.5 μL of annealed sgRNA 2 at 200 nM concentration per reaction. As described below, to achieve triple *SOXF* knockdown, we transduced CRISPRi

hPSCs with a *mU6-SOX18-sgRNA*; *EF1A-HygroR* plasmid, and subsequently transduced them with this dual *SOX7/SOX17* sgRNA construct.

All plasmids were transformed into Mix & Go! *E. coli* Competent Cells (Zymo Research) and purified using the Wizard Plus SV Minipreps DNA Purification System (Promega).

CRISPRi knockdown in hPSCs

CRISPRi plasmids were packaged into VSV-G pseudotyped lentiviruses in HEK293T/17 cells using a 3rd generation lentiviral packaging system as described previously[48]. 24 hours prior to transfection, HEK293T/17 cells were seeded at a density of 105,000 cells/cm^2 in 6-well plates coated with 0.01% poly-L-Lysine. Each well was transfected with 1.39 μg pMDL plasmid + 0.78 μg VSV-G plasmid + 0.53 μg pREV plasmid + 11.3 μL FuGENE HD transfection reagent + 2.1 μg sgRNA plasmid in Opti-MEM medium. 18 hours post-transfection, media was changed to DMEM with 10% FBS. Supernatant was collected 42 hours post-transfection, filtered with 0.45 μM polyethersulfone filter, and stored at -80 °C until used for transduction.

In parallel, H1 hPSCs were engineered to constitutively express CRISPR interference (CRISPRi) machinery, namely nuclease-dead Cas9 (dCas9) fused to the transcriptional repressor ZIM3 KRAB[72]. Validation of these CRISPRi-expressing H1 hPSCs will be described in a forthcoming manuscript.

To transduce them with lentiviruses carrying single sgRNAs, H1 CRISPRi hPSCs were dissociated with Accutase, and then re-seeded at 150,000 cells/well in 6-well plates in StemFlex (Thermo Fisher, A3349401) supplemented with 1% penicillin/streptomycin and 1 μM thiazovivin and lentivirus-containing supernatant (15 μL for *SOX17*-sgRNA, 10 μL for *SOX7*-sgRNA, and 50 μL for *SOX18*-sgRNA lentiviruses, respectively). After reaching 80% confluency after 2-3 days, cells were re-plated using Accutase and cultured in StemFlex with Hygromycin B (Thermo Fisher, 10687010) at 50-100 μg/ml concentration for at least 3 weeks. 1 μM thiazovivin was supplemented for the first day of selection.

To generate the triple *SOXF* knockdown line, first we transduced H1 CRISPRi hPSCs with the *mU6-SOX18-sgRNA*; *EF1A-HygroR* lentivirus and performed hygromycin selection for at least 3 weeks, as described above. These H1 CRISPRi *SOX18*-sgRNA hPSCs were then dissociated with Accutase, re-seeded at 200,000 cells/well in 12-well plates in StemFlex supplemented with 1% penicillin/streptomycin, 50 μg/ml Hygromycin B and 1 μM thiazovivin and transduced with 50 μl of lentivirus-containing supernatant. The following day, media was replaced with StemFlex supplemented with 1% penicillin/streptomycin and 50 μg/ml Hygromycin. After 3 days, cells were re-plated using Accutase and selected with 1 μg/ml Puromycin (Thermo Fisher, A1113803) and 50 μg/mL Hygromycin B for at least 3 days. 1 μM thiazovivin was supplemented for the first day of antibiotic selection.

H1 CRISPRi hPSC lines were then further differentiated into artery and vein ECs using the V1 protocol described above. qPCR and scRNAseq was performed on control and *SOXF* knockdown artery ECs, primed ECs, and vein ECs as described above.

Construction of *SOX17-2A-mPlum; NR2F2-2A-GFP* reporter hPSC line

The Cas9/AAV6 knock-in strategy[139] was used to generate a double *SOX17-2A-mPlum*; *NR2F2-2A-GFP* knock-in reporter hPSC line. We previously generated single *SOX17-2A-mPlum* or *NR2F2-2A-GFP* reporter hPSC lines[48]. In this study, starting from a *NR2F2-2A-GFP* reporter line (clone 10), we used our previously-described approach to knock-in a *2A-mPlum* cassette downstream of the endogenous *SOX17* gene such that the *SOX17* stop codon was removed[48].

Regulatory and ethical oversight

All animal procedures at Stanford University were approved by Stanford's Administrative Panel on Laboratory Animal Care (APLAC). All animal procedures at Oxford University were approved by a local ethical review committee at Oxford University and licensed by the UK Home Office. All hPSC experiments at Stanford University were approved by Stanford's Stem Cell Research Oversight (SCRO) committee.

QUANTIFICATION AND STATISTICAL ANALYSES

Statistical tests are specified in the figure legend accompanying each experimental result. Unpaired t-tests were used to test for statistical significance in qPCR and ELISA data, and P values were reported. Wilcoxon rank sum tests were performed to test for statistical significance in module score differences quantified by scRNAseq, and P values were reported. The following convention was used in figures: not significant (n.s., $P>0.05$), $*P<0.05$, and $**P<0.01$.

Q-values were reported for differentially-expressed genes detected by bulk-population RNA-seq, differentially-accessible genomic loci detected by Omni-ATAC-seq, and transcription factor motif enrichments detected by Omni-ATAC-seq. The q-value represents the P value adjusted for multiple hypothesis testing using the Benjamini-Hochberg method to control the false discovery rate.

Quantification of mouse embryo images was performed as described in the section "Quantification of mouse embryo images".

All *in vitro* experiments entailed two or more technical replicates, defined as two or more wells of cells subjected to the same culture conditions in the same experiment. The only exceptions were experiments that entailed single-cell RNA-seq and LEGENDScreen profiling; such profiling was performed on one technical replicate. For all qPCR experiments, two separate measurements were made of each technical replicate (i.e., for each well of cultured cells, qPCR for each given gene was performed twice in the same 384-well qPCR plate).

For mouse embryo experiments, the following numbers of embryos were analyzed per experiment, and one representative image is shown:

- **E8.5-E8.75 mouse embryos, stained for Erg and Sox17** (**Fig. 4C-D**, **Fig. S8C**): N=5 embryos
- **E9.5 mouse embryos, stained for Erg and Sox17** (**Fig. 4E**): N=3 embryos
- **E9.5 mouse embryos, stained for *Aplnr*, Erg, and Sox17** (**Fig. 4F**): N=9 embryos
- **E12.5 mouse embryos, *Sox17-Cre; Aplnr-DreER* intersectional lineage tracing** (**Fig. 4H**): N=3 embryos

2011 • **E12.5 mouse embryos, *Sox17-Cre* lineage tracing** (**Fig. S8E**): N=3 embryos
2012 • **E9.75 mouse embryos, *Aplnr-CreER* lineage tracing** (**Fig. S8F**): N=7 embryos
2013 • **Quantification of E9.5 mouse embryos, *Aplnr-CreER* lineage tracing** (**Fig. S8G**): N=4
2014 embryos. Images of 6 regions of interest (ROI) were quantified per embryo.
2015
2016 No sample size calculation, randomization, or investigator blinding were performed.

SUPPLEMENTAL INFORMATION

Table S1: **Single-cell RNA-seq of hPSC-derived cell-types**
Differentially expressed genes that distinguish day 0 hPSCs ("H1"), day 1 mid primitive streak ("d1ps"), day 2 lateral mesoderm ("d2dlm"), day 3 artery ECs ("d3aus"), day 3 primed ECs ("d3pvus"), day 4 vein ECs ("d4vus"), and day 3-4 mesenchyme ("mes1"). Gene expression was measured by single-cell RNA-seq.

Table S2: **Bulk RNA-seq of hPSC-derived cell-types**
Count matrix of gene expression of day-0 hPSCs ("H1"), day-1 mid primitive streak ("D1_PS"), day-2 lateral mesoderm ("D2_DLM"), day 3 artery ECs, CD144+ FACS-purified ("D3_Artery"), day 3 primed ECs, CD144+ FACS-purified ("D3_Prevein"), and day 4 vein ECs, CD144+ FACS-purified ("D4_Vein"). Gene expression was measured by bulk-population RNA-seq, and is quantified in $\log_2$ counts per million (CPM) units.

Table S3: **Cell-surface marker screening of hPSC-derived cell-types**
For each hPSC-derived cell-type, the number indicates the percentage of cells that expressed a given cell-surface marker by high-throughput flow cytometry. Day 3 artery EC and day 4 vein EC populations were pre-gated on the CD144+ endothelial subset before quantifying the expression of other cell-surface markers. This was done to avoid being confounded by CD144- non-endothelial cells in the culture.

Table S4: **Single-cell RNA-seq of E9.5 mouse embryo ECs *in vivo***
Differentially expressed genes that distinguish 4 different types of ECs identified in the E9.5 mouse embryo. Raw single-cell RNA-sequencing data were obtained from a previous study[140]. We provisionally annotated the following clusters based on expression of characteristic marker genes: 0 (vein-like 1 ECs), 1 (other ECs), 2 (artery ECs), and 3 (vein-like 2 ECs).

Table S5: **Single-cell RNA-seq of gene expression changes upon triple *SOXF* knockdown during artery EC differentiation**
Genes that are differentially expressed between artery ECs generated from control vs. *SOX7/SOX17/SOX18*-CRISPRi hPSCs. Positive fold change indicates higher expression in control artery ECs, and $\log_2$ fold change is shown. Gene expression was measured by single-cell RNA-seq.

Table S6: **Single-cell RNA-seq of gene expression changes upon triple *SOXF* knockdown during vein EC differentiation**
Genes that are differentially expressed between vein ECs generated from control vs. *SOX7/SOX17/SOX18*-CRISPRi hPSCs. Positive fold change indicates higher expression in control vein ECs, and $\log_2$ fold change is shown. Gene expression was measured by single-cell RNA-seq.

Table S7: **Differentially accessible genomic elements in hPSC-derived artery vs. vein ECs, as assayed by OmniATAC-seq**
Differentially accessible genomic elements in hPSC-derived artery vs. vein ECs, as assayed by OmniATAC-seq. Fold change indicates accessibility in vein ECs relative to artery ECs (i.e.,

2063 a positive value indicates that a genomic element is more accessible in vein ECs, whereas a
2064 negative value indicates greater chromatin accessibility in artery ECs). The gene closest to
2065 each genomic element is also indicated.
2066
2067 **Table S8**: **Transcription factors enriched in artery-specific regulatory elements, as**
2068 **assayed by OmniATAC-seq**
2069 Transcription factor motifs enriched in artery-specific regulatory elements, as determined by
2070 HOMER. Motifs are named in accordance with the most closely matching motif in the HOMER
2071 dataset. P value indicates the statistical significance of motif enrichment in artery-specific
2072 regulatory elements, relative to vein-specific regulatory elements.
2073
2074 **Table S9**: **Transcription factors enriched in vein-specific regulatory elements, as**
2075 **assayed by OmniATAC-seq**
2076 Transcription factor motifs enriched in vein-specific regulatory elements, as determined by
2077 HOMER. Motifs are named in accordance with the most closely matching motif in the HOMER
2078 dataset. P value indicates the statistical significance of motif enrichment in vein-specific
2079 regulatory elements, relative to artery-specific regulatory elements.