Transcriptional dynamics of pluripotent stem cell-derived endothelial cell differentiation revealed by single-cell RNA sequencing

ESC European Society of Cardiology

European Heart Journal (2019) **00**, 1–14
doi:10.1093/eurheartj/ehz351

Transcriptional dynamics of pluripotent stem cell-derived endothelial cell differentiation revealed by single-cell RNA sequencing

Ian R. McCracken[1], Richard S. Taylor[1,2], Fatma O. Kok[1†], Fernando de la Cuesta[1], Ross Dobie[2], Beth E.P. Henderson[2], Joanne C. Mountford[3], Axelle Caudrillier[1], Neil C. Henderson[2], Chris P. Ponting[4], and Andrew H. Baker [1*]

[1]Centre for Cardiovascular Science, University of Edinburgh, 47 Little France Crescent, Edinburgh EH16 4TJ, UK; [2]Centre for Inflammation Research, University of Edinburgh, Edinburgh EH16 4TJ, UK; [3]Scottish National Blood Transfusion Service, Edinburgh EH14 4BE, UK; and [4]MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK

Received 24 September 2018; revised 13 March 2019; editorial decision 9 May 2019; accepted 14 May 2019

Aims Pluripotent stem cell-derived endothelial cell products possess therapeutic potential in ischaemic vascular disease. However, the factors that drive endothelial differentiation from pluripotency and cellular specification are largely unknown. The aims of this study were to use single-cell RNA sequencing (scRNA-seq) to map the transcriptional landscape and cellular dynamics of directed differentiation of human embryonic stem cell-derived endothelial cells (hESC-EC) and to compare these cells to mature endothelial cells from diverse vascular beds.

Methods and results A highly efficient directed 8-day differentiation protocol was used to generate a hESC-derived endothelial cell product (hESC-ECP), in which 66% of cells co-expressed CD31 and CD144. We observed largely homogeneous hESC and mesodermal populations at Days 0 and 4, respectively, followed by a rapid emergence of distinct endothelial and mesenchymal populations. Pseudotime trajectory identified transcriptional signatures of endothelial commitment and maturation during the differentiation process. Concordance in transcriptional signatures was verified by scRNA-seq analysis using both a second hESC line RC11, and an alternative hESC-EC differentiation protocol. In total, 105 727 cells were subjected to scRNA-seq analysis. Global transcriptional comparison revealed a transcriptional architecture of hESC-EC that differs from freshly isolated and cultured human endothelial cells and from organ-specific endothelial cells.

Conclusion A transcriptional bifurcation into endothelial and mesenchymal lineages was identified, as well as novel transcriptional signatures underpinning commitment and maturation. The transcriptional architecture of hESC-ECP was distinct from mature and foetal human EC.

Keywords Endothelial differentiation • Embryonic stem cell • Single-cell RNA sequencing • Pseudotime • RNA velocity • Vascular regeneration

* Corresponding author. Tel: +44 131 24 26728, Email: andy.baker@ed.ac.uk

† Present address. Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

Translational perspective

Ischaemic conditions contribute substantially to global mortality and morbidity. Human embryonic stem cell-derived endothelial cells (hESC-EC) offer a promising strategy to promote therapeutic angiogenesis in tissue ischaemia. Our good manufacturing practice (GMP) compliant protocol for the generation of hESC-EC promotes therapeutic angiogenesis *in vivo*. However, knowledge of the transcriptional dynamics of hESC-EC differentiation is currently limited. Here, we apply single-cell RNA sequencing to two hESC differentiation protocols. We characterize the transcriptional dynamics of hESC-EC, important for informing how to drive controlled angiogenesis, endothelial cell maturation, and specification. We also evaluate the cellular heterogeneity at each stage using a hESC-EC differentiation protocol that is relevant to clinical cell therapy.

Introduction

Cardiovascular disease remains the most common cause of mortality globally despite improved pharmacological and interventional options.[1] For example, current interventions for critical limb ischaemia are unsuccessful, with 43% of patients requiring limb amputation within 5 years.[2] Similarly, for myocardial infarction 25% of patients still progress to heart failure owing to the overburdening of the remaining viable myocardium.[3]

Consequently, using stem-cell therapies to promote vascular regeneration is a promising strategy for treating ischaemic disease. However, despite studies yielding positive results in small animal models of ischaemia, the efficacy of using such cells in human trials has been disappointing, providing limited benefit.[4]

While many protocols have tested autologous cells, a number of studies have recently assessed the use of pluripotent cells in relevant animal models[5] and more recently clinically.[6] As such, derivation and production of therapeutic cells from pluripotent sources are required using an efficient and GMP compliant protocol. Transplantation of pluripotent stem cell-derived endothelial cells (PSC-EC) has been demonstrated to promote therapeutic angiogenesis in a range of small animal models of ischaemia.[7,8] However, differentiation protocols are limited by their low efficiency resulting in substantial and uncharacterized heterogeneity within the final cell population.[9]

We recently reported a rapid and efficient protocol[7] for production of a regenerative human embryonic stem cell-derived endothelial cell product (hESC-ECP), adapted from the protocol of Patsch *et al.*[10] The protocol consistently yields CD31+/CD144+ cells comprising ≈60% of the population, with the remaining cells expressing pericyte and mesenchymal markers and, crucially, exhibiting no residual pluripotency.[7] When injected into ischaemic muscle, a robust pro-angiogenic response was evident across a range of mouse models.[7] Despite the expression of mature endothelial cell (EC) markers by cells from the hESC-ECP, it is likely that these cells remain transcriptionally immature and are non-specified further into arterial, venous, or lymphatic lineages.[7]

Comprehensive characterization of the transcriptional dynamics throughout differentiation is a crucial step towards understanding the key molecular regulators and mechanisms governing commitment and maturation of EC. Additionally, defining the temporal changes in pluripotent and angiogenic-related genes may provide insight into the safety and functional efficacy of these cells. Recent studies of cardiac and liver differentiation have demonstrated that longitudinal application of single-cell RNA sequencing (scRNA-seq) permits tracking of transcriptional changes throughout differentiation at single-cell resolution.[11,12]

A recent study used scRNA-seq to attempt to understand the directed differentiation of induced pluripotent stem cells-derived EC, yet the final cell population was highly heterogeneous, consisting of fewer than 10% EC and co-populated by cardiac-like and hepatic-like cells.[9] Development of droplet-based scRNA-seq platforms has enabled the sequencing of thousands of cells in parallel, and the rigorous unbiased modelling of cellular heterogeneity.[13] Consequently, application of these platforms for characterizing human embryonic stem cell-derived endothelial cells (hESC-EC) has clear advantages over previous studies that surveyed fewer than 100 cells.[14]

In this study, we sought to use scRNA-seq to dissect the cellular heterogeneity and transcriptional dynamics of hESC-EC differentiation and maturation, utilizing alternative hESC lines and two different differentiation protocols. Maturity and specification of hESC-EC following their initial commitment were assessed throughout the process by comparison to foetal and adult endothelial cells from various sources. We also aim to evaluate the clinical suitability of our hESC-ECP by comparison to the products of an alternative differentiation.

Methods

Human embryonic stem cell-derived endothelial cell differentiation

H9 and RC11 hESC lines were differentiated to hESC-ECP using our previously reported protocol.[7] Human ESC lines were used in accordance with the UK Stem Cell Bank Steering Committee guidelines (Project Approvals SCS11-51 and SCSC17-26). Briefly, hESC were plated on a fibronectin matrix at Day 0 (d0). At d1, lateral mesoderm was induced with GSK3 inhibitor (CHIR99021) (7 μM) and BMP4 (25 ng/mL) added to N2B27/Neurobasal/DMEM:F12 media. This was followed by endothelial cell induction at d4 with Forskolin (2 μM) and vascular endothelial growth factor (VEGF) (200 ng/mL) in StemPro34 media. In the final step, cells were replated and cultured until d8 or d12 with human AB serum (1%) and VEGF (50 ng/mL) in EGM-2 media. Cells were taken for scRNA-seq at d0 (H9 only), d4, d6, d8, and d12 (H9 only).

In the alternative differentiation protocol H9 hESC were differentiated to hESC-EC as described by Zhang *et al.*[15] In summary, hESC were plated as single cells on a vitronectin matrix and cultured in E8 media supplemented with BMP4 (5 ng/mL), Activin A (25 ng/mL), and CHIR99021(1 μM). From d2 to d6, media was changed to E5 media supplemented with FGF (100 ng/mL), VEGF (50 ng/mL), SB431542 (10 μM), RESV (resveratrol) (5 μM), and L690 (10 μM). Cells were harvested for scRNA-seq at d2, d4, and d6.

Single-cell RNA sequencing library construction

Following determination of their number and viability, dissociated cells were loaded into the 10X Chromium Controller. Library construction was performed using the Single Cell 3' Reagent Kit. To summarize, in reaction vesicles (gel beads in emulsion, GEMs), cells were lysed and barcoded oligonucleotides reverse transcribed before clean-up and cDNA amplification. The Chromium Single-Cell 3' Library Kit was then used to generate indexed sequencing libraries. Library sequencing was performed using either Illumina HiSeq 4000 or Illumina NovaSeq 6000 platforms. Recovery rates ranged from between 3778 and 10 053 cells, with the number of reads per cell ranging between 35 630 and 119 015 (Supplementary material online, *Table S1*).

Single-cell RNA sequencing data analysis

Raw reads were processed using the 10X CellRanger pipeline to align these to the reference transcriptome (GRCh38) and to generate gene-cell count matrices. Initial QC and clustering was performed with the aid of Seurat version 2.2.1.[16] In an attempt to remove dead or falsely identified cells, as well as doublets, cells either expressing fewer than 250 genes, having a UMI count greater than 10 000, or having greater than 15% of reads mapping to mitochondrial genes were excluded from further analysis. Following data normalization and scaling to remove unwanted sources of variation[16] principal component analysis (PCA) was conducted using genes with highly variable expression. Seurat graph-based clustering[16] was then applied followed by tSNE allowing the visualization of identified clusters in tSNE plots. Differential gene expression (DGE) analysis was then used to identify significantly differentially expressed genes within each cluster using the Wilcoxon test for significance (adjusted P value <0.05) and a $\log_e$(FC) greater than 0.25. Signature scores were calculated by taking the mean of the scaled and centred expression values across multiple signature genes (Supplementary material online, *Methods*).

Pseudotime trajectories of differentiation were generated using the Monocle R Package v.2.6.4.[17] Pseudotime analysis proceeds on the basis that cells undergo biological processes in an asynchronous manner, and thus that cells can be ordered along a calculated trajectory to infer the transcriptional changes throughout the process. Genes that were most differentially expressed ($q < 10^{-40}$) in identified clusters were used to assign pseudotime values to individual cells. Differentially expressed genes across pseudotime were identified using the differentialGeneTest command in Monocle.[17] RNA velocity of individual cells was calculated using the Velocyto R package (V0.6.0).[18] RNA velocity analysis determined the fraction of spliced-to-unspliced reads to predict the future transcriptional state of individual cells. All sequencing data have been deposited in the Gene Expression Omnibus (GEO; GSE131736).

All further experimental and analysis details are included in the Supplementary material online, *Methods*.

Results

Longitudinal Single-cell RNA sequencing analysis reveals the emergence of committed endothelial and mesenchymal lineages from a homogeneous mesodermal population

Human embryonic stem cells (H9) underwent an 8-day differentiation protocol[7] via a mesodermal stage to form a hESC-EC product (*Figure* 1A). At Day 8, two-thirds (65.7%) of the cells were $CD31^+$/$CD144^+$ with fewer than 0.05% co-expressing pluripotent markers SSEA4/TRA-181 as shown by flow cytometry (*Figure 1B*). Single-cell transcriptomics data were obtained from 21 369 cells harvested at four time points across the differentiation protocol using droplet-based microfluidic technology (*Figure 1A*).

Cellular heterogeneity was characterized at each time point using PCA. The starting hESC population (d0) was largely homogeneous: 99.7% and 95.9% of cells expressed pluripotent markers *POU5F1* (*OCT4*) and *SOX2*, respectively (Supplementary material online, *Figure S1A*).

At Day 4 (d4), 97.4% of cells expressed lateral mesoderm marker *HAND1* (*Figure 1C*). The paucity of cells expressing early and late endothelial markers [*KDR* (5.4%) and *PECAM1* (0.16%), respectively] demonstrated that this population was not pre-determined to an endothelial lineage (*Figure 1C*). In cluster d4-B, reduced expression of *MIXL1* (*Figure 1C*) coupled to a higher expression of bone marrow stromal cell marker *BST2* (Supplementary material online, *Figure S1B*), suggested pre-disposition to a mesenchymal fate.[19]

In contrast, cells at Day 6 (d6) and Day 8 (d8) were dichotomized into two divergent populations (*Figure 1C*). The first was characterized by expression of endothelial markers (*KDR* and *PECAM1*), and the second by mesenchymal markers such as *ACTA2* (*Figure 1C*). Residual heterogeneity within both populations at either d6 or d8 reflects cell cycle variation, indicated by *MKI67* expression (*Figure 1C*). Cluster d8-F (3.2%) likely represented a doublet cluster owing to the expression of both mesenchymal and endothelial markers, whilst lacking unique markers (Supplementary material online, *Figure S1D*).

By d8, fewer than 0.5% of cells expressed either *POU5F1* or *SOX2* with no cells co-expressing both markers (Supplementary material online, *Figure S2*), confirming the absence of residual pluripotency. Mesodermal markers *MIXL1* and *HAND1* were expressed in 12% and 14% of d8 cells, respectively (Supplementary material online, *Figure S2*) confirming a non-mesodermal population.

At Day 8, hESC-EC and mesenchymal populations expressed angiogenesis associated genes such as *TGFB1*, *HIF1A*, and *FLT1* suggesting contribution to the pro-angiogenic effect (Supplementary material online, *Figure S3*). Other pro-angiogenic markers, such as *TIE1* and *ANGPT2*, were highly expressed in the endothelial lineage whilst *TGFB2* and *FGF1* were primarily expressed within mesenchymal clusters.

To investigate phenotypical changes beyond d8, we repeated the differentiation and processed an additional 13 657 H9 cells at d8 and d12. Separation of endothelial and mesenchymal populations at Day 8 was replicated in both rounds of differentiation as well as in a previous replicate ($n = 3$) (Supplementary material online, *Figure S4*). Distinct endothelial and mesenchymal populations were still evident at Day 12 (*Figure 1D*). The proportion of endothelial cells in cycle decreased from 49.6% at d8 to 23.1% at d12 (D12-B). No cluster representing cycling cells was identified in the d12 mesenchymal population.

Genes that were differentially expressed in cluster D8-E (such as *ENPP2*, *TIE1*, and *CLDN5*) showed elevated expression in hESC-EC at d12 in comparison to at d8 (*Figure 1E*).[20–22] This is consistent with the observed increase in cell confluency (Supplementary material online, *Figure S5*). Furthermore, Gene Ontology (GO) term enrichment analysis revealed extracellular matrix organization to be a significantly

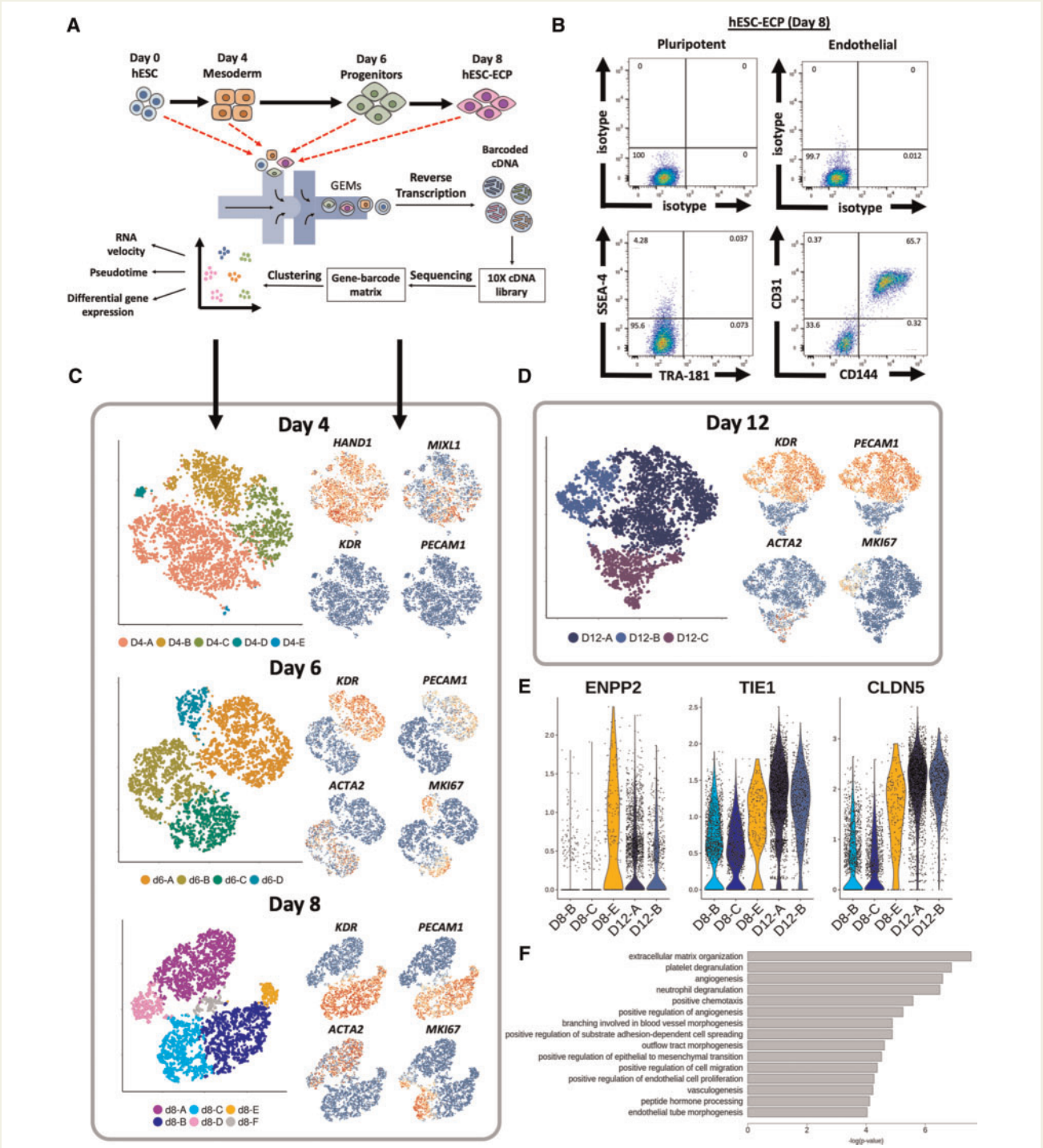


Figure 1 Single-cell RNA sequencing analysis of human embryonic stem cells-derived endothelial cell product differentiation at individual time points. (*A*) Schematic of human embryonic stem cells-derived endothelial cell product differentiation and single-cell RNA sequencing pipeline. (*B*) Representative flow cytometric analysis of cells taken at Day 8 (human embryonic stem cells-derived endothelial cell product). Cells stained for endothelial (CD31 and CD144) (right panels), pluripotent markers (TRA-181 and SSEA-4) (left panels), and their corresponding isotype controls. (*C*) tSNE plots, constructed separately, using data from cells sampled at Days 4, 6, and 8 for single-cell RNA sequencing. Feature plots (right) show key marker gene expression for each population. (*D*) tSNE plot constructed from d12 data from a separate round of H9 differentiation. (*E*) Violin plots showing expression (ln(UMI + 1)) of angiogenic and cell contact related genes in d8 and d12 human embryonic stem cell-derived endothelial cells. (*F*) Gene Ontology term analysis conducted using genes significantly differentially expressed within d12 human embryonic stem cell-derived endothelial cells in comparison to d8 human embryonic stem cell-derived endothelial cells.

enriched term (*Figure* 1F). At each time point, hESC-ECs predominantly expressed arterial markers as opposed to venous or lymphatic ECs (Supplementary material online, *Figure S6*).

Longitudinal single-cell RNA sequencing analysis reveals rapid endothelial differentiation and subsequent maturation following a bifurcation point

In order to characterize the transcriptional changes throughout differentiation, data from all time points were combined and normalized (see Methods section) prior to conducting PCA and data visualization (tSNE). Cells from d0 and d4 each formed a separate cluster, whereas d6 and d8 cells separate into endothelial or mesenchymal populations (*Figure 2A*). To trace transcriptional dynamics associated with each identified cell type, expression values of pluripotent, mesodermal, mesenchymal, and endothelial markers were used to calculate a score for each of the four signatures for each individual cell (see Supplementary material online, *Methods*). As expected, 100% of d0 cells were positive for the pluripotent signature (*Figure 2B* and *C*), whereas virtually no d4 cells were positive for this signature. Similarly, d4 cells were mainly positive for the mesodermal signature (*Figure 2B* and *C*) with enriched expression of *APLNR* (Supplementary material online, *Figure S7*). Increased endothelial and mesenchymal signature scores for d6 and d8 within their respective populations suggested that both cell types have matured further following initial specification (*Figure 2B* and *C*).

The concept of RNA velocity can be used to predict future transcriptional states, taking advantage of changes in read fractions in exonic vs. intronic sequence.[18] RNA velocities in *Figure 2E* are represented by the size and direction of arrows. The paucity of vectors with large magnitude and coherent direction associated with d0 and d4 cells is indicative of their equilibrium state (*Figure 2E*). Vectors associated with either endothelial or mesenchymal cell clusters indicated the direction of cell type maturity. The larger vectors for the d8 mesenchymal cell cluster, compared with the d8 endothelial cluster, is consistent with the latter reaching a steadier equilibrium state following initial maturation (*Figure 2E*).

Early (d6) and late (d8) EC and mesenchymal populations exhibit high transcriptional similarity (*Figure 2D*). Nevertheless, higher expression of transcription factor (TF) *SNAI2* was observed in the early (d6) mesenchymal population in comparison to later mesenchymal population (d8) (Supplementary material online, *Figure S8*). Similarly, higher expression of genes such as *SOX7* and *GJA4* at d6 defines these as markers of early endothelial specification (Supplementary material online, *Figure S9A* and *B*). Conversely, higher expression of *GNG11, ESM1*, and *ANGPT2* at d8 defines these as later markers of endothelial maturation (Supplementary material online, *Figure S9A* and *B*). RT-qPCR analysis of several marker genes utilizing RNA from CD144+ and CD144- populations at d6 and d8 demonstrated high concordance with scRNA-seq findings (Supplementary material online, *Figure S10*). A complete list of significantly differentially expressed genes for all compared datasets is provided in Supplementary material online, *File S1*.

To ensure findings were not limited to H9 cells, an alternative hESC line, the RC11 line, was used to reproduce findings. Cells were sampled for scRNA-seq at Days 4, 6, and 8 (total of 21 972 cells) and the identical analysis applied. Strikingly, we identified equivalent mesodermal, mesenchymal, and endothelial populations in both RC11- and H9-derived cells (*Figure 2F*). A small additional cluster was identified defined by differential expression epithelial markers including *EPCAM*, *CDH1*, and *RAB25* (Supplementary material online, *Figure S11*). This likely reflects an epithelial-to-mesenchymal transition during mesoderm formation.[23]

Pseudotime analysis is a complementary approach used to determine the path and progress of individual cells undergoing differentiation.[17] The resultant trajectory for H9 hESC-ECP differentiation (*Figure 3A*) indicated a bifurcation point giving rise to distinct endothelial and mesenchymal branches, consistent with the divergence represented in *Figure 2A*. The majority of d4 cells had pseudotime values close to zero, implying they had yet to commence endothelial or mesenchymal differentiation (Supplementary material online, *Figure S12*). Pseudotime analysis using RC11 cells reproduced an essentially identical bifurcation point as using H9 (*Figure 3B*). Increasing expression of known endothelial markers (*Figure 3C*) and TF (*Figure 3D*) defined the endothelial branch. Similarly, up-regulation of mesenchymal markers and TF (*Figure 3C and D*) identified the mesenchymal branch (*Figure 3A*). Expression of the same markers and TF during the RC11 differentiation traced highly similar pseudotime trajectories (*Figure 3C* and *D*).

Single-cell RNA sequencing analysis of an alternative human embryonic stem cell-derived endothelial cell differentiation protocol reveals substantial heterogeneity

A comparable longitudinal scRNA-seq data set was acquired using an alternative hESC-EC differentiation protocol, namely the five-factor protocol described by Zhang *et al.*[15] This allowed us to compare its cellular heterogeneity and transcriptional dynamics with those from our 8-day differentiation protocol. Using the Zhang *et al.* 6-day protocol, differentiated H9 cells yielded 51.8% CD31+/CD144+ cells by Day 6 (Supplementary material online, *Figure S13*). Cells were sampled for scRNA-seq at a mesodermal stage (d2) and two endothelial stages (d4 and d6). We sequenced the transcriptomes of 27 225 cells. Unsupervised clustering revealed the presence of 12 individual clusters (*Figure 4A* and *B*).

Within the d2 population, a large (22%) epithelial cluster was identified, characterized by expression of markers including *EPCAM, RAB25*, and *CLDN4* (*Figure 4C* and *D*). The remaining d2 cells were characterized by expression of mesodermal markers *HAND1* and *BMP4*. Interestingly, d2 mesoderm was also positive for mesendoderm markers such as *GSC* and *EOMES* (*Figure 4C* and *D*). Expression of a pluripotent marker, *NANOG*, was predominantly localized to the epithelial cluster (Supplementary material online, *Figure S14A*).

By d4, 37% of cells localized to a distinct endothelial population defined by expression of classical endothelial markers (*Figure 4A–D*). The two-remaining major d4 clusters were defined by expression of mesodermal markers (*Figure 4A–D*). In contrast to d2, d4 mesoderm was predominantly negative for expression of mesendoderm markers (*Figure 4D*). The 18.5% reduction of *EOMES*+ cells in cluster 'mesoderm 2', in comparison to cluster 'mesoderm 1' likely identifies this as the later of the two d4 mesoderm populations. Interestingly, a subset

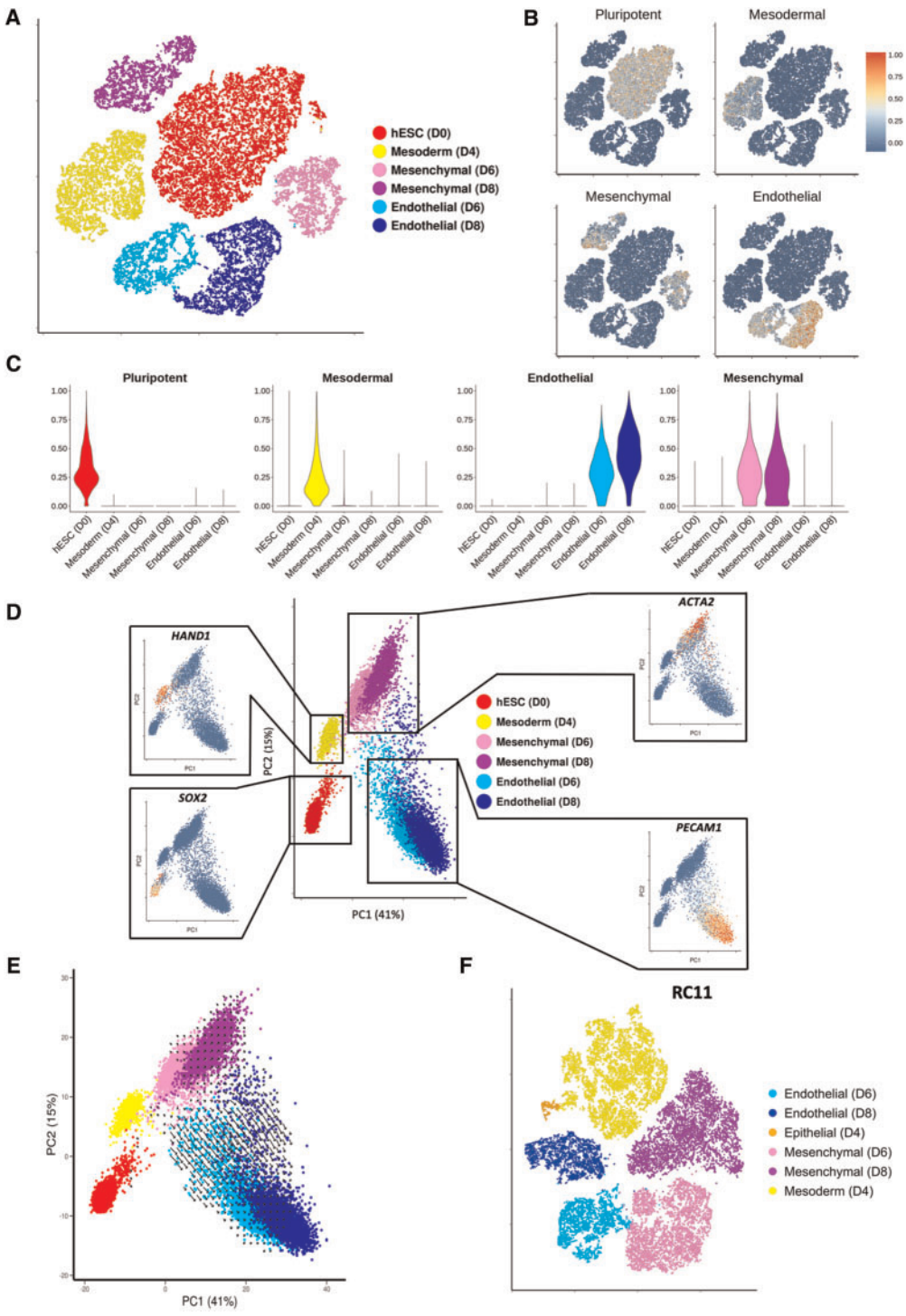


Figure 2 Single-cell RNA sequencing analysis across all time points mapping the transcriptional changes and predicted future direction of cells undergoing human embryonic stem cells-derived endothelial cell product differentiation. (*A*) tSNE plot constructed using data from 21 369 cells taken at Day 0, 4, 6, and 8 during H9 to human embryonic stem cells-derived endothelial cell product differentiation. Cluster identity determined by the expression of known pluripotent (*SOX2*), mesodermal (*HAND1*), mesenchymal (*ACTA2*), and endothelial markers (*PECAM1*). (*B*) Cell signature scores shown on tSNE plot from *A*. (*C*) Cell signature scores within identified clusters from *A*. (*D*) Principal component analysis plot containing cells taken at the four time points during H9 differentiation. Feature plots show key marker gene expression. (*E*) RNA velocities visualized on the principal component analysis plot from *D*. (*F*) tSNE plot using data from d4, d6, and d8 during differentiation of RC11.

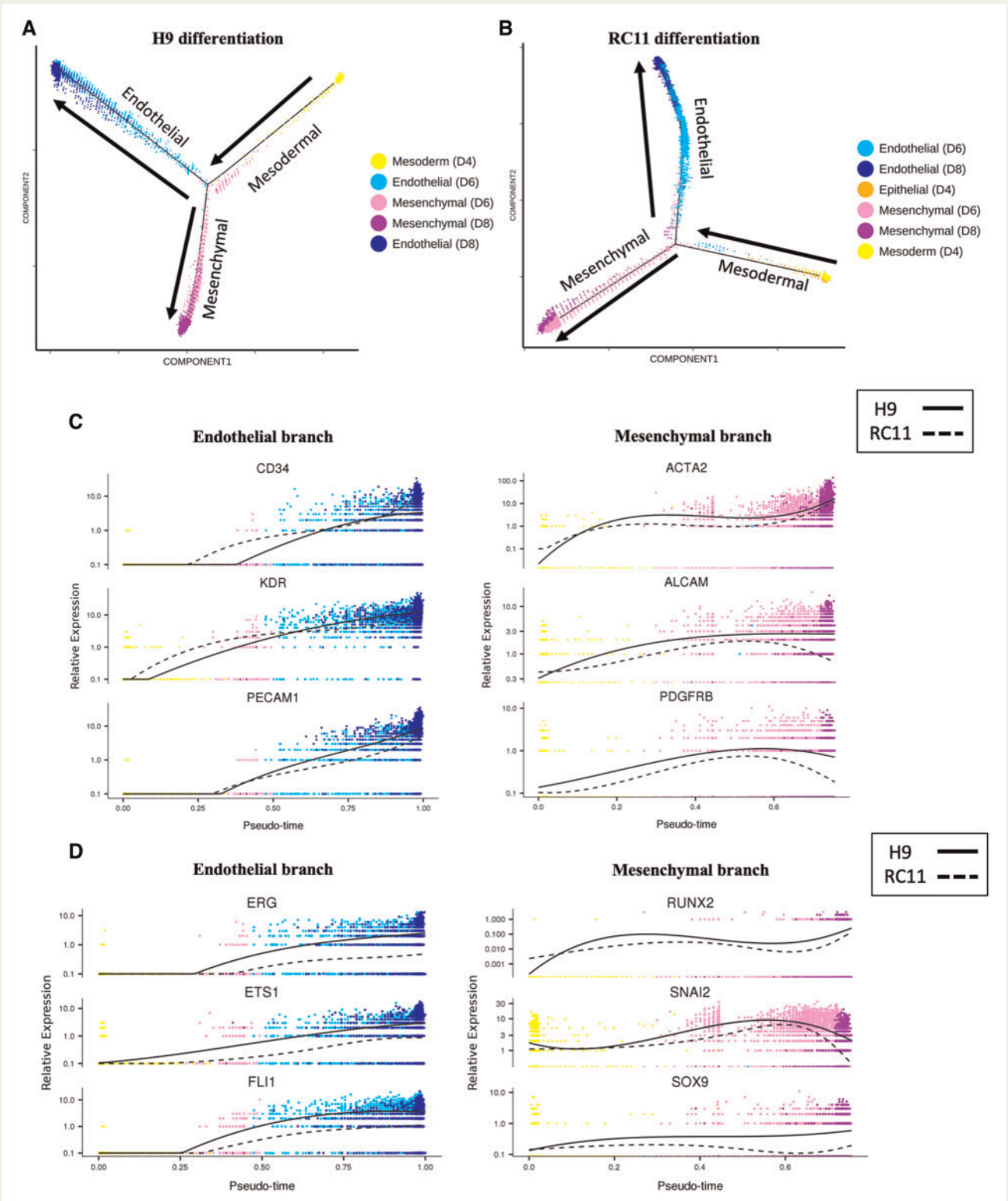


Figure 3 Pseudotime trajectories of human embryonic stem cells-derived endothelial cell product differentiation. Pseudotime trajectories containing cells from d4, d6, and d8 during (*A*) H9 differentiation and (*B*) RC11 differentiation. Arrows show direction of pseudotime across trajectories. Expression of known endothelial and mesenchymal (*C*) markers and (*D*) transcription factors in their respective branches for each of the two starting ES cells (H9 and RC11).

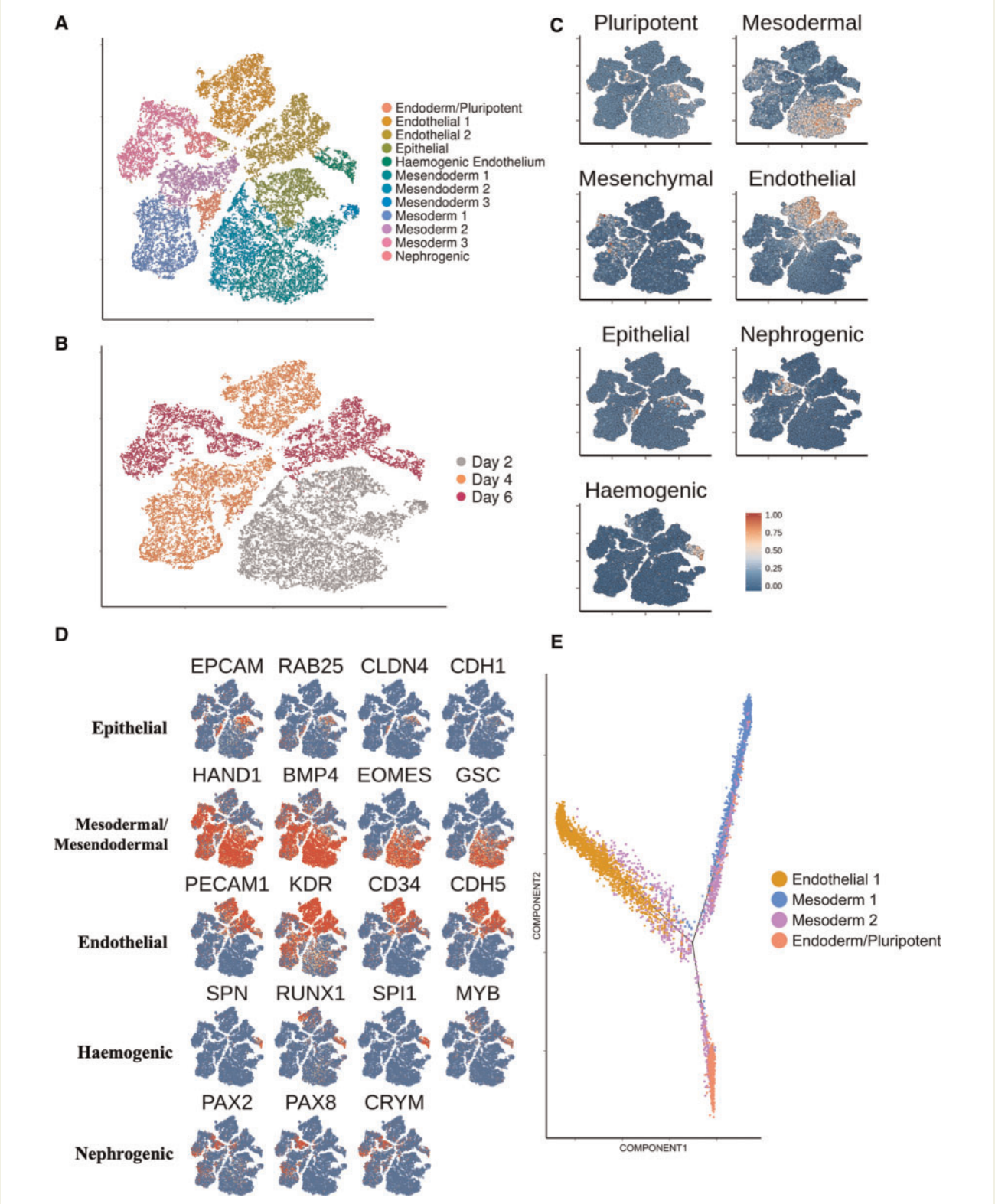


Figure 4 Longitudinal single-cell RNA sequencing analysis of an alternative human embryonic stem cell-derived endothelial cells differentiation protocol. (*A*) tSNE plot showing data of human embryonic stem cell-derived endothelial cells differentiation using an alternative protocol. Cells taken for single-cell RNA sequencing at d2, d4, and d6. (*B*) tSNE plot of alternative differentiation protocol, categorizing cells according to their original data set. (*C*) Signature scores shown on tSNE plot from *A*. (*D*) Key marker gene expression used to characterize identified clusters. (*E*) Pseudotime trajectory containing cells from d4 clusters of the alternative differentiation protocol.

of cluster 'mesoderm 2' expressed both *CD34* and *KDR* but not *PDGFRA* suggesting a more specialized mesoderm with restricted specification to a hematoendothelial lineage[24] (Supplementary material online, *Figure S14B*). We annotated a small cluster of d4 cells that differentially express endoderm markers, including *FOXA2* and *CLDN6* as well as pluripotent marker *POU5F1*, as an early endoderm cluster with residual pluripotency (Supplementary material online, *Figure S14C*).

At d6, a comparable *HAND1*+ mesoderm population was present (*Figure 4A–D*). The large d6 endothelial cluster accounted for 43% of d6 cells (*Figure 4A–D*). About 10.3% of d6 cells localized to a small cluster exclusively expressing various known haemogenic markers including *SPN1*, *SPI1*, and *RUNX1*[25] (*Figure 4C* and *D*). Furthermore, a small d6 cluster (10.7% of d6 cells) was annotated as being a nephrogenic population, defined by expression of *PAX2, PAX8*, and *CRYM*[26] (*Figure 4C* and *D*).

We next used pseudotime analysis choosing cells at d4 over d6 so as to compare early vs. late mesodermal populations. As expected, the resultant trajectory (*Figure 4E*) revealed cells from the 'mesoderm 1' cluster preceding those from 'mesoderm 2'. Unexpectantly, the bifurcation point giving rise to the endodermal population appeared to emerge from the later, 'mesoderm 2' population. Endothelial and mesodermal cells populated the start of the endothelial branch, thus suggesting that endothelial commitment from mesoderm was still in progress at d4 (*Figure 4E*).

Pseudotime analysis characterizes transcription factor expression throughout endothelial differentiation

We sought to characterize the expression of transcription factors over these trajectories. Using a comprehensive list of over 1600 likely human TF[27] we identified 144 that were significantly differentially expressed over pseudotime in both RC11 and H9 differentiations. Of these, several are known to be implicated in endothelial differentiation (*ERG, ETS1*, *SOX7*,and *ELF1*[28]) and these all resided within Cluster 5 (containing 37 genes) of genes differentially expressed across pseudotime (*Figure 5A*).

The expression profile of the 37 Cluster 5 TF (*Figure 5A*) was then visualized along the endothelial branch of the trajectory from the five-factor protocol (*Figure 5B*). Expression dynamics for the majority of TF showed substantial resemblance to those observed in the five-factor protocol (*Figure 5C*; Supplementary material online, *Figure S15*). In addition, several TF with no clear reported link to endothelial differentiation also followed a similar expression profile, and showed remarkable concordance across cells and protocols (*Figure 5C*).

Human embryonic stem cell-derived endothelial cells remain transcriptionally distinct from mature and adult endothelial cells despite undergoing additional maturation following endothelial commitment

We then sought to evaluate the maturity of hESC-EC produced by comparison of scRNA-seq data of foetal endothelial cells and mature endothelial cells with that from H9-derived cells. First, we combined hESC-EC data from d6, d8, and d12 datasets with data obtained from mature primary cultured ECs from each of the arterial, lymphatic, or venous vascular beds. Principal component analysis revealed that hESC-EC at d6, d8, and d12 cluster separately from mature EC (*Figure 6A*). Interestingly, d12 hESC-EC clustered separately from their d6 and d8 counterparts, but not visibly closer to any given mature EC. Large proportions of genes were expressed within hESC-EC but not in mature EC (d8: 7.2%; d12: 7.3%; *Figure 6B*). Differential gene expression analysis between d8 and d12 hESC-EC revealed increasing expression levels of several ECM related genes, such as *COL4A1* and *FN1* (Supplementary material online, *Figure S16A*). Furthermore, between d8 and d12 hESC-EC an increase in endothelial signature score was observed, suggesting of further maturation beyond d8 (Supplementary material online, *Figure S16B*).

We then collected all publicly available 10X Chromium scRNA-seq data sets including endothelial cell data from freshly isolated human organs (Supplementary material online, *Table S3*). Differential gene expression analysis of combined endothelial clusters (*Figure 6C*) revealed key transcriptional markers for each organ-specific endothelial phenotype (Supplementary material online, *File S2*). Visualization of signature scores across all clusters demonstrated their high specificity to their corresponding cluster (*Figure 6D*). Neither the d8 nor d12 hESC-EC cluster was positive for any organ-specific signature score, thus indicating their likely unspecified nature. Furthermore, correlations between transcriptomes of endothelial cell types were higher among the d6, d8, and d12 cells, than these cells were to foetal (kidney) and mature endothelial cells (*Figure 6E*).

Discussion

In the present study, we performed the first longitudinal droplet-based scRNA-seq analysis of our highly efficient hESC-EC differentiation protocol.[7] In two different hESC lines, endothelial differentiation from a homogeneous lateral mesoderm population resulted in a bifurcation leading to the emergence of non-specified endothelial and mesenchymal populations. Each population was capable of undergoing further maturation. We performed comparable scRNA-seq analysis of an alternative hESC-EC differentiation protocol, revealing substantially more heterogeneity during differentiation, including the emergence of multiple non-endothelial populations by the protocol endpoint. Pseudotime analysis comparing different protocols and starting hESC lines uncovered robust transcriptional signatures of endothelial differentiation. Finally, we compared the hESC-EC transcriptional identity to those of primary cultured endothelial cells and freshly isolated foetal and adult endothelial cells. Despite hESC-EC appearing to undergo additional maturation following their initial commitment, they remained transcriptionally distinct from foetal and mature endothelial cell types, with no indication of commitment to an organ-specific endothelial phenotype.

In contrast to earlier time points, at d6 we observed the emergence of two distinct populations expressing either mesenchymal or endothelial markers, a dichotomy that was confirmed by pseudotime trajectory analysis. Recent studies have reported the existence of a common mesenchyme and endothelial precursor, termed the mesenchymoangioblast (MB) derived from APLNR+/KDR+/PDGFRA+ mesoderm.[29] MB-derived mesenchymal cells were demonstrated to emerge from an endothelial origin by endothelial-to-mesenchymal

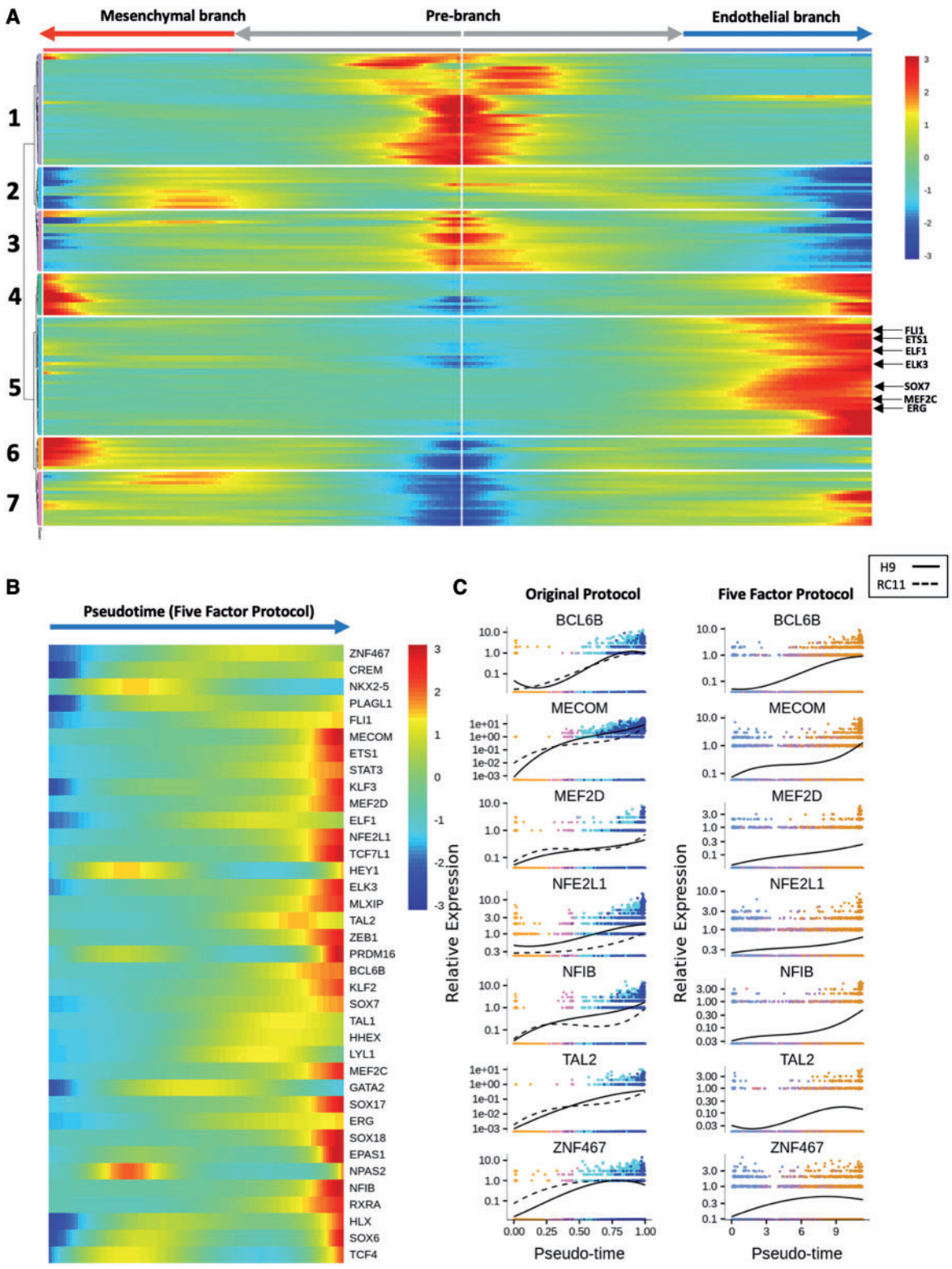


Figure 5 Pseudotime analysis of transcription factor expression during human embryonic stem cell-derived endothelial cells differentiation. (*A*) Differential gene expression variance over pseudotime of H9 differentiation within pseudotime trajectory branches. Genes used are TF that are differentially expressed in both H9 and RC11-derived trajectories. (*B*) Gene expression variance of TF taken from Cluster 5 of heatmap from *A* in the endothelial branch of the trajectory from the five-factor protocol. (*C*) Spline plots showing the expression of novel TF from *B* across pseudotime in RC11 and H9 human embryonic stem cell-derived endothelial cells differentiation using the original protocol, as well as for the five-factor protocol.

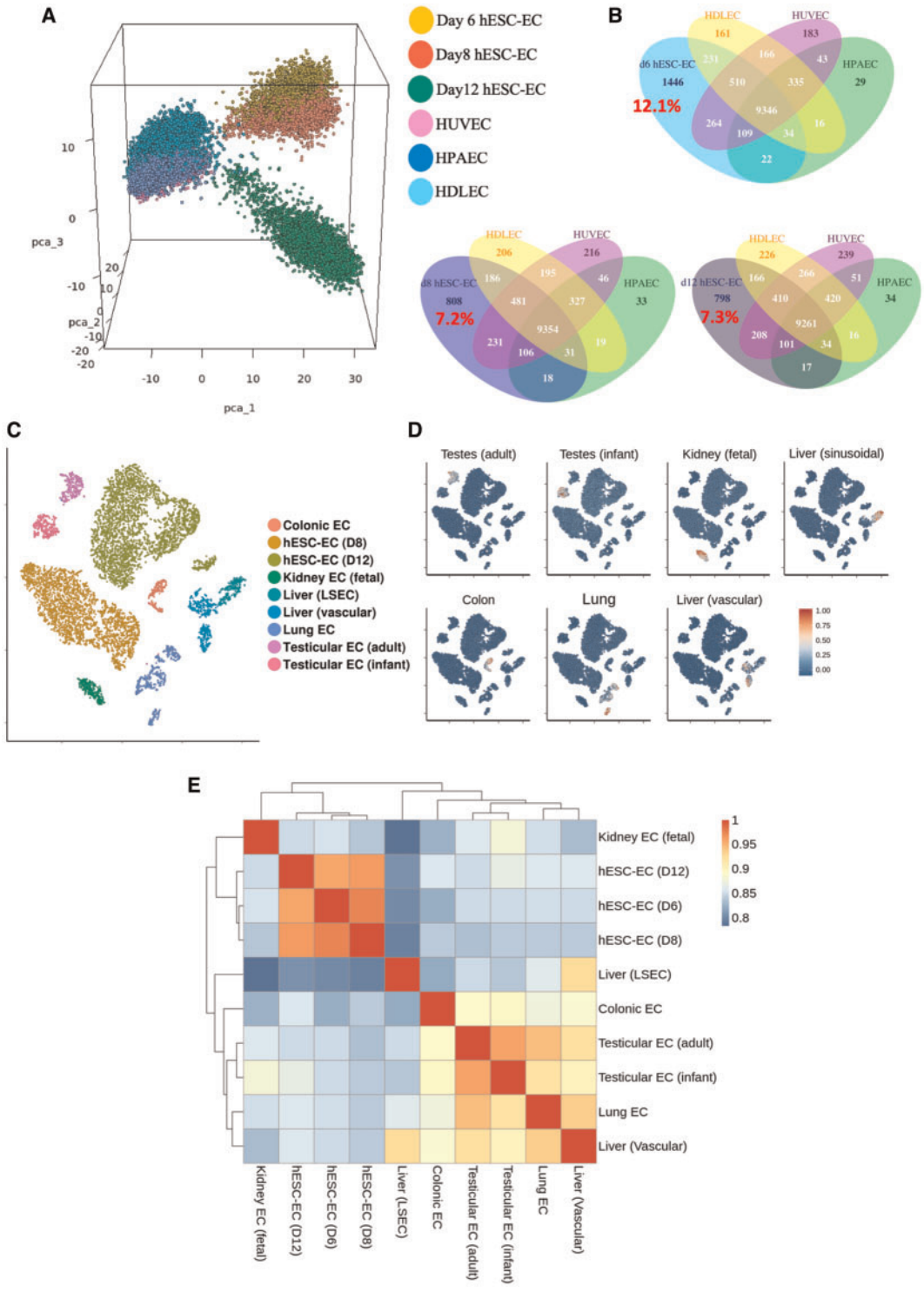
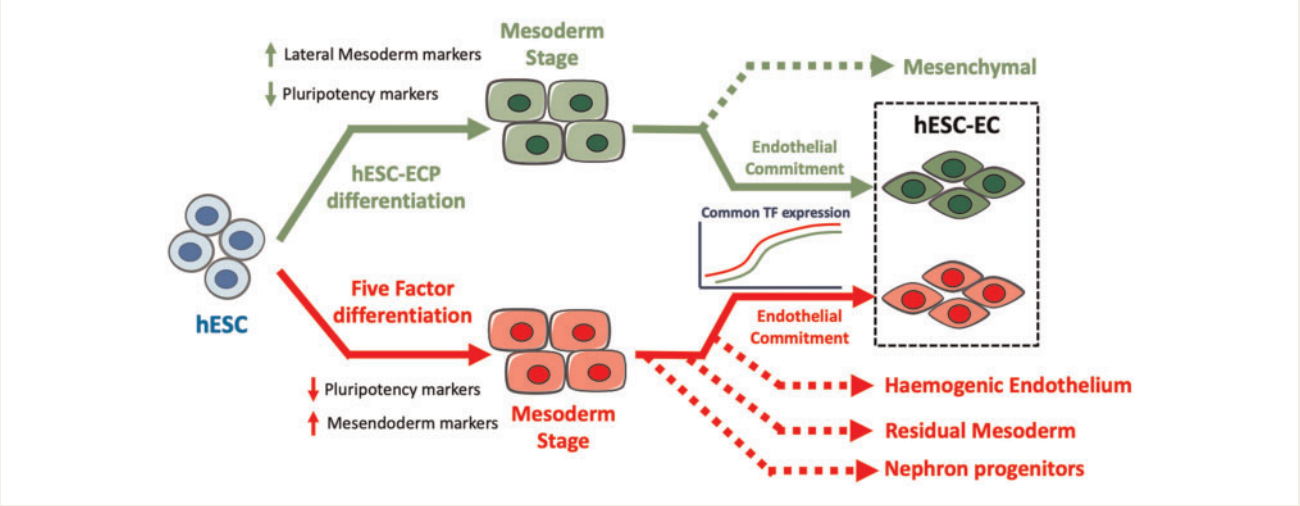


Figure 6 Transcriptomic comparison of human embryonic stem cell-derived endothelial cells to mature endothelial cells. (*A*) Three-dimensional principal component analysis plot of human embryonic stem cell-derived endothelial cells from d6, d8, and d12 datasets alongside mature endothelial cell: human dermal lymphatic endothelial cells, human umbilical vein endothelial cells, and human pulmonary artery endothelial cells. (*B*) Venn diagrams of number of genes expressed in d6, d8, and d12 human embryonic stem cell-derived endothelial cells and the overlap of these genes with each type of mature endothelial endothelial cell. Percentage of uniquely expressed human embryonic stem cell-derived endothelial cell genes shown in red. (*C*) tSNE plot showing clustering of endothelial cell extracted from 10X datasets of human organs. (*D*) Signature scores for each endothelial cell type shown in *C*. (*E*) Heatmap with hierarchical clustering comparing transcriptional status of each cell type by Spearman's rank correlation coefficient.



Take home figure Proposed model for the differentiation of the embryonic stem cells-derived endothelial cells via two differentiation methods. Human embryonic stem cells undergo highly efficient differentiation using the human embryonic stem cells-derived endothelial cell product differentiation method, forming a homogeneous lateral mesoderm population. Following endothelial induction, a bifurcation point exists, giving rise to early endothelial and mesenchymal populations. Conversely, in the five-factor protocol, human embryonic stem cells undergo differentiation to an earlier mesendoderm population. Subsequent endothelial induction drives the mesendoderm to give rise to multiple unwanted cell types including nephron progenitors and haemogenic endothelium. Despite clear differences in cellular heterogeneity, transcriptional dynamics of human embryonic stem cells-derived endothelial cell differentiation has substantial commonalities between differentiation methods.

transition (EndMT) in the absence of VEGF.[30] However, other mesenchymal progenitors, called mesopheres, were found to develop in the presence of VEGF and have shared phenotypic and transcriptomic properties with MBs.[31] The observed expression of the EndMT-associated TF *SNAI2* in the early mesenchymal population may suggest induction of EndMT in the presence of VEGF.[32] However, here, although the Day 4 mesodermal population selectively expressed *APLNR*, it was found not to express *KDR* or *PDGFRA*, thus suggesting that mesenchymal and endothelial commitment is more likely to occur independently from a common mesoderm progenitor.[24]

In contrast to our own 8-day differentiation protocol,[7] analysis of the alternative, five-factor differentiation protocol[15] revealed considerably more heterogeneity, with an early mesendoderm population identified by primitive streak markers and residual pluripotency (*Take home figure*). Interestingly, the model proposed by Evseenko *et al.*[24] shows early mesoderm can then undergo further specification to KDR+/CD34+/PDGFRA- mesoderm which has limited specification potential, forming only hematoendothelial lineages. This model may explain the unexpected haemogenic endothelial cluster in the d6 population. The transcriptional signatures identified around the haemogenic endothelial cluster demonstrated high concordance with scRNA-seq analysis of an iPSC model of endothelial-to-hematopoietic transition (EHT).[25]

We also identified expression patterns of several known and novel TFs in all three hESC-EC pseudotime trajectories. For example, the expression profile of *BCL6B*, which encodes a transcription factor shown to be expressed within the endothelium of the developing mouse kidney,[33] showed remarkable similarity across all trajectories. This suggests a potential role in endothelial differentiation and/or maturation, making this a strong candidate for further experimental validation. Additional cues such as shear stress are likely to be required to drive EC maturation.[34] Our initial attempts to subject d8 hESC-EC to shear stress resulted in detachment of cells (data not shown), possibly due to cells not being at a sufficient level of maturation to lay down the extracellular matrix necessary for effective adhesion.

In addition, recent cell tracking studies revealed that the majority of cells are lost soon after transplantation suggesting a paracrine dependence *in vivo*.[7,35] Crucially with scRNA-seq analysis, several angiogenic factors such as *FLT1* and *HIF1A*, were highly expressed across both endothelial and mesenchymal populations in our d8 hESC-ECP. Furthermore, the non-teratogenic nature of the hESC-ECP was further demonstrated by the complete absence of cells co-expressing *SOX2* and *POU5F1*.

In conclusion, we have used a droplet-based scRNA-seq platform to map the transcriptional changes throughout the differentiation of hESCs to pluripotent stem cell-derived EC across different cell lines and differentiation protocols. This provides novel insight into early endothelial commitment and maturation and will aid in the refinement of protocols that drive endothelial specification in a lineage or organ-specific nature. We also provide essential evidence of the transcriptional state and heterogeneity by Day 8 of our protocol, and we are developing a first-in-man cell therapy approach using Day 8 differentiated cells.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Acknowledgements

The authors thank G. Aitchison and K. Newton for their technical assistance. Flow cytometry data were generated with support from the QMRI Flow Cytometry and cell sorting facility, University of Edinburgh. Sequencing data was produced with assistance from Edinburgh Genomics.

Funding

This work was supported by the Medical Research Council [MRC Precision Medicine Doctoral Training Programme to I.R.M. and both the MRC Discovery Award and Programme grant (MC_PC_15075) and MRC Programme: Computational and Disease Genomics (MC_UU_00007/15) to C.P.P.], the Wellcome Trust [Wellcome Trust Senior Research Fellowship in Clinical Science (ref. 1037[49]) to N.C.H.], the European Research Council [Advanced Grant VASCMIR (RE7644) to A.H.B.], and the British Heart Foundation [BHF CVR grant (RM/17/3/33381) and BHF Chair of Translational Cardiovascular Sciences to A.H.B.].

Conflict of interest: none declared.

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;**385**:9963117–9963171.
2. Howard DP, Banerjee A, Fairhead JF, Hands L, Silver LE, Rothwell PM. Population-based study of incidence, risk factors, outcome, and prognosis of ischemic peripheral arterial events: implications for prevention. *Circulation* 2015;**132**:1805–1815.
3. Minicucci MF, Azevedo PS, Polegato BF, Paiva SA, Zornoff LA. Heart failure after myocardial infarction: clinical implications and treatment. *Clin Cardiol* 2011;**34**:410–414.
4. Tompkins BA, Balkan W, Winkler J, Gyongyosi M, Goliasch G, Fernandez-Aviles F, Hare JM. Preclinical studies of stem cell therapy for heart disease. *Circ Res* 2018;**122**:1006–1020.
5. Liu YW, Chen B, Yang X, Fugate JA, Kalucki FA, Futakuchi-Tsuchida A, Couture L, Vogel KW, Astley CA, Baldessari A, Ogle J, Don CW, Steinberg ZL, Seslar SP, Tuck SA, Tsuchida H, Naumova AV, Dupras SK, Lyu MS, Lee J, Hailey DW, Reinecke H, Pabon L, Fryer BH, MacLellan WR, Thies RS, Murry CE. Human embryonic stem cell-derived cardiomyocytes restore function in infarcted hearts of non-human primates. *Nat Biotechnol* 2018;**36**:597–605.
6. Menasche P, Vanneaux V, Hagege A, Bel A, Cholley B, Cacciapuoti I, Parouchev A, Benhamouda N, Tachdjian G, Tosca L, Trouvin JH, Fabreguettes JR, Bellamy V, Guillemain R, Suberbielle Boissel C, Tartour E, Desnos M, Larghero J. Human embryonic stem cell-derived cardiac progenitors for severe heart failure treatment: first clinical case report. *Eur Heart J* 2015;**36**:2011–2017.
7. MacAskill MG, Saif J, Condie A, Jansen MA, MacGillivray TJ, Tavares AAS, Fleisinger L, Spencer HL, Besnier M, Martin E, Biglino G, Newby DE, Hadoke PWF, Mountford JC, Emanueli C, Baker AH. Robust revascularization in models of limb ischemia using a clinically translatable human stem cell-derived endothelial cell product. *Mol Ther* 2018;**26**:1669–1684.
8. Rufaihah AJ, Huang NF, Jame S, Lee JC, Nguyen HN, Byers B, De A, Okogbaa J, Rollins M, Reijo-Pera R, Gambhir SS, Cooke JP. Endothelial cells derived from human iPSCS increase capillary density and improve perfusion in a mouse model of peripheral arterial disease. *Arterioscler Thromb Vasc Biol* 2011;**31**:e72–e79.
9. Paik DT, Tian L, Lee J, Sayed N, Chen IY, Rhee S, Rhee JW, Kim Y, Wirka RC, Buikema JW, Wu SM, Red-Horse K, Quertermous T, Wu JC. Large-scale single-cell RNA-seq reveals molecular signatures of heterogeneous populations of human induced pluripotent stem cell-derived endothelial cells. *Circ Res* 2018;**123**:443–450.
10. Patsch C, Challet-Meylan L, Thoma EC, Urich E, Heckel T, O'Sullivan JF, Grainger SJ, Kapp FG, Sun L, Christensen K, Xia Y, Florido MHC, He W, Pan W, Prummer M, Warren CR, Jakob-Roetne R, Certa U, Jagasia R, Freskgård P-O, Adatto I, Kling D, Huang P, Zon LI, Chaikof EL, Gerszten RE, Graf M, Iacone R, Cowan CA. Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. *Nat Cell Biol* 2015;**17**:994–1003.
11. Camp JG, Sekine K, Gerber T, Loeffler-Wirth H, Binder H, Gac M, Kanton S, Kageyama J, Damm G, Seehofer D, Belicova L, Bickle M, Barsacchi R, Okuda R, Yoshizawa E, Kimura M, Ayabe H, Taniguchi H, Takebe T, Treutlein B. Multilineage communication regulates human liver bud development from pluripotency. *Nature* 2017;**546**:533–538.
12. Loh KM, Chen A, Koh PW, Deng TZ, Sinha R, Tsai JM, Barkal AA, Shen KY, Jain R, Morganti RM, Shyh-Chang N, Fernhoff NB, George BM, Wernig G, Salomon REA, Chen Z, Vogel H, Epstein JA, Kundaje A, Talbot WS, Beachy PA, Ang LT, Weissman IL. Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types. *Cell* 2016;**166**:451–467.
13. Haque A, Engel J, Teichmann SA, Lonnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017;**9**:75.
14. Ohta R, Niwa A, Taniguchi Y, Suzuki NM, Toga J, Yagi E, Saiki N, Nishinaka-Arai Y, Okada C, Watanabe A, Nakahata T, Sekiguchi K, Saito MK. Laminin-guided highly efficient endothelial commitment from human pluripotent stem cells. *Sci Rep* 2016;**6**:35680.
15. Zhang J, Chu LF, Hou Z, Schwartz MP, Hacker T, Vickerman V, Swanson S, Leng N, Nguyen BK, Elwell A, Bolin J, Brown ME, Stewart R, Burlingham WJ, Murphy WL, Thomson JA. Functional characterization of human pluripotent stem cell-derived arterial endothelial cells. *Proc Natl Acad Sci USA* 2017;**114**:E6072–E6078.
16. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–420.
17. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;**32**:381–386.
18. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lonnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundstrom E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. RNA velocity of single cells. *Nature* 2018;**560**:494–498.
19. Roson-Burgo B, Sanchez-Guijo F, Del Canizo C, De Las Rivas J. Insights into the human mesenchymal stromal/stem cell identity through integrative transcriptomic profiling. *BMC Genomics* 2016;**17**:944.
20. Nam SW, Clair T, Kim YS, McMarlin A, Schiffmann E, Liotta LA, Stracke ML. Autotaxin (NPP-2), a metastasis-enhancing motogen, is an angiogenic factor. *Cancer Res* 2001;**61**:6938–6944.
21. Yuan L, Le Bras A, Sacharidou A, Itagaki K, Zhan Y, Kondo M, Carman CV, Davis GE, Aird WC, Oettgen P. ETS-related gene (ERG) controls endothelial cell permeability via transcriptional regulation of the claudin 5 (CLDN5) gene. *J Biol Chem* 2012;**287**:6582–6591.
22. Chan B, Yuan HT, Ananth Karumanchi S, Sukhatme VP. Receptor tyrosine kinase Tie-1 overexpression in endothelial cells upregulates adhesion molecules. *Biochem Biophys Res Commun* 2008;**371**:475–479.
23. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 2014;**15**:178–196.
24. Evseenko D, Zhu Y, Schenke-Layland K, Kuo J, Latour B, Ge S, Scholes J, Dravid G, Li X, MacLellan WR, Crooks GM. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc Natl Acad Sci U S A* 2010;**107**:13742–13747.
25. Guibentif C, Ronn RE, Boiers C, Lang S, Saxena S, Soneji S, Enver T, Karlsson G, Woods NB. Single-cell analysis identifies distinct stages of human endothelial-to-hematopoietic transition. *Cell Rep* 2017;**19**:10–19.
26. Narlis M, Grote D, Gaitan Y, Boualia SK, Bouchard M. Pax2 and pax8 regulate branching morphogenesis and nephron differentiation in the developing kidney. *J Am Soc Nephrol* 2007;**18**:1121–1129.
27. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The human transcription factors. *Cell* 2018;**172**:650–665.
28. De Val S, Black BL. Transcriptional control of endothelial cell development. *Dev Cell* 2009;**16**:180–195.
29. Slukvin, II, Kumar A. The mesenchymoangioblast, mesodermal precursor for mesenchymal and endothelial cells. *Cell Mol Life Sci* 2018;**75**:3507–3520.
30. Vodyanik MA, Yu J, Zhang X, Tian S, Stewart R, Thomson JA, Slukvin II. A mesoderm-derived precursor for mesenchymal stem and endothelial cells. *Cell Stem Cell* 2010;**7**:718–729.
31. Gertow K, Hirst CE, Yu QC, Ng ES, Pereira LA, Davis RP, Stanley EG, Elefanty AG. WNT3A promotes hematopoietic or mesenchymal differentiation from hESCs depending on the time of exposure. *Stem Cell Reports* 2013;**1**:53–65.

32. Cooley BC, Nevado J, Mellad J, Yang D, St Hilaire C, Negro A, Fang F, Chen G, San H, Walts AD, Schwartzbeck RL, Taylor B, Lanzer JD, Wragg A, Elagha A, Beltran LE, Berry C, Feil R, Virmani R, Ladich E, Kovacic JC, Boehm M. TGF-beta signaling mediates endothelial-to-mesenchymal transition (EndMT) during vein graft remodeling. *Sci Transl Med* 2014;**6**:227ra34.
33. Yu J, Valerius MT, Duah M, Staser K, Hansard JK, Guo JJ, McMahon J, Vaughan J, Faria D, Georgas K, Rumballe B, Ren Q, Krautzberger AM, Junker JP, Thiagarajan RD, Machanick P, Gray PA, van Oudenaarden A, Rowitch DH, Stiles CD, Ma Q, Grimmond SM, Bailey TL, Little MH, McMahon AP. Identification of molecular compartments and genetic circuitry in the developing mammalian kidney. *Development* 2012;**139**:1863–1873.
34. Atkins GB, Jain MK. Role of Kruppel-like transcription factors in endothelial biology. *Circ Res* 2007;**100**:1686–1695.
35. Wysoczynski M, Khan A, Bolli R. New paradigms in cell therapy: repeated dosing, intravenous delivery, immunomodulatory actions, and new cell types. *Circ Res* 2018;**123**:138–158.