Modeling vascular diseases using human induced pluripotent stem cells

Cao, X.

Chapter 4

Transcriptional Dynamics During Segregation of Endothelial and Myocardial Lineages from Cardiac Mesoderm

Xu Cao[1], Maria Mircea[2], Gopala Krishna Yakala[1], Francijna E. van den Hil[1], Hailiang Mei[3], Konstantinos Anastassiadis[4], Christine L. Mummery[1], Stefan Semrau[2] and Valeria V. Orlova[1, *]

[1]Department of Anatomy and Embryology, Einthovenweg 20, 2333ZC Leiden, The Netherlands
[2]Leiden Institute of Physics, Leiden University, 2333 RA, Leiden, The Netherlands
[3]Sequencing analysis support core, Leiden University Medical Center, Leiden, 2333ZA, The Netherlands
[4]Stem Cell Engineering, Biotechnology Center, Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, Dresden, 01307, Germany
*Corresponding author

Manuscript in preparation

ABSTRACT

We previously showed that two major cellular components of the human heart, cardiac endothelial cells and cardiomyocytes, could be derived simultaneously from human induced pluripotent stem cells (hiPSCs). *In vivo* studies in mice have shown these lineages derive from a common MESP1+ mesoderm progenitor. Although this has been investigated using whole transcriptome sequencing, the transcriptional control and dynamics of their segregation from cardiac mesoderm are not yet fully understood. In addition, the same transgenic approach cannot be used in human. Here, we used bulk and single cell RNA sequencing (RNAseq) to investigate EC and cardiomyocytes co-differentiation in hiPSC. We confirmed segregation of these two cardiac lineages from common cardiac mesoderm precursors but more importantly, revealed a critical role of transient expression of the transcription factor ETV2 for endothelial fate specification and showed unexpectedly that functional cardiomyocytes could also originate from ETV2+ progenitors.

INTRODUCTION

The ability to obtain different cardiac lineages derived from human induced pluripotent stem cells (hiPSCs) can facilitate studies of human heart development and modeling of cardiovascular diseases *in vitro*. Cardiomyocytes and endothelial cells (ECs) are two major cell types in the heart and their crosstalk through physical interactions and paracrine, autocrine and endocrine factors (Tirziu et al., 2010) has allowed the generation of physiologically relevant cardiac (micro)tissues, organoids and engineered heart tissues *in vitro* (Caspi et al., 2007; Giacomelli et al., 2017; Goldfracht et al., 2020; Masumoto et al., 2016; Mills et al., 2017; Narmoneva et al., 2004; Ravenscroft et al., 2016; Stevens et al., 2009; Tulloch et al., 2011).

Cardiomyocytes and ECs originate from *Mesp1*+ progenitors specified during gastrulation. In mice, these cells are first located in primitive streak and then migrate toward lateral plate mesoderm at ~E6.5 (Devine et al., 2014; Lescroart et al., 2014; Meilhac et al., 2014; Saga et al., 1999). It is still controversial when the segregation of cardiomyocytes and ECs from common progenitors actually occurs. scRNAseq of mouse *Mesp1*+ progenitors collected at E6.75 and E7.25 showed that these cells were already segregated into distinct cardiovascular lineages, including cardiomyocytes and ECs (Lescroart et al., 2018). However, some studies showed multipotential progenitors were still present in lateral plate mesoderm which originates in the primitive streak and expresses *Flk-1* (Ema et al., 2006; Garry and Olson, 2006). These cells were initially recognized as multipotent cardiac progenitor cells (CPCs) (Tam et al., 1997). However, in human heart development, there is limited knowledge on the timing and transcriptional control for segregation of cardiac endothelial and myocardial lineages from multipotent CPCs.

Studies in mouse and chick showed heart mainly develops from two sources of progenitors (Buckingham et al., 2005; Vincent and Buckingham, 2010): the first heart field (FHF) in the cardiac crescent contributes to the primitive heart tube which serves as a scaffold into which second heart field (SHF) cells can migrate before heart chamber morphogenesis. It had already been shown that cells from SHF were patterned before migration in to give rise to different parts of heart (Galli et al., 2008; Zaffran et al., 2004). CPCs from FHF and SHF could be distinguished by expression of ISL1, which was specifically expressed in SHF (Cai et al., 2003). NKX2-5 expressing CPCs in both FHF and SHF from E7.5 to E7.75 could contribute to both the cardiomyocytes and ECs in the heart (Paffett-Lugassy et al., 2013). As a direct target of NKX2-5, ETV2 was found to be expressed in all ECs but not myocardium by E8.5 (Ferdous et al., 2009). ETV2 was required for the development of endothelial and hematopoietic lineages and directly targets *TAL1*, *GATA2*, *LMO2*, *TEK*, *NOTCH1*, *NOTCH4*, *CDH5* (De Val et al., 2008; Ferdous et al., 2009; Lee et al., 2008; Liu et al., 2015). Overexpression of ETV2 itself could convert human fibroblasts into endothelial-like cells (Morita et al., 2015). In mouse embryonic stem cells (ESCs), ETV2 could initiate the hemangiogenic specification in a threshold-dependent manner instructed by VEGF signaling (Zhao et al., 2017). Later, ETV2 expression was also found to direct the segregation of hemangioblasts and smooth muscle cells (SMCs) in mouse ESCs (Zhao and Choi, 2019). However, it is still unclear whether ETV2 also plays a role in the segregation of ECs and cardiomyocytes from CPCs during human heart development.

Previously we found that *MESP1*+ progenitors derived from human ESCs could give rise to cardiomyocytes, ECs and SMCs (Hartogh et al., 2015). *Flk-1*+ cardiovascular progenitors generated from ESCs resembled CPCs *in vivo* (Moretti et al., 2006). Recently we developed a co-differentiation system for ECs and cardiomyocytes from hiPSCs through a common cardiac mesoderm (CM) stage (Giacomelli et al., 2017). ECs thus derived showed a cardiac specific identity, evidenced by high expression of cardiac markers *MEOX2*, *GATA4*, *GATA6* and *ISL1*. The developmental path and transcriptional dynamics during differentiation of endothelial and myocardial lineages in this protocol have, however, not yet been characterized.

Droplet-based single cell RNA sequencing (scRNAseq) great facilitates the study of global transcriptional dynamics at single cell resolution. Paik et al. performed scRNAseq analysis of hiPSC-derived ECs (hiPSC-ECs) which were less than 10% of the total population among cells expressing the cardiac maker *TNNT2.* The developmental process of EC and cardiac lineages as such were not further studied (Paik et al., 2018). In another scRNAseq study on hiPSC-ECs from a different differentiationmethod, pseudotime analysis of ECs collected at different time points showed that endothelial and mesenchymal lineages had a common developmental origin in mesoderm cells (McCracken et al., 2019), although the identity and differentiation potential of these

mesenchymal cells was not characterized.

Here, we performed both scRNAseq and bulk RNAseq on cells from our co-differentiation system. scRNAseq analysis identified CM as the common developmental origin of cardiomyocytes and ECs. *ETV2* expression was observed principally in the endothelial lineage but also in a subpopulation of the myocardial lineage. An ETV2^{mCherry} hiPSC reporter line was generated and used to study real-time expression dynamics and the role of ETV2 during endothelial and myocardial lineage segregation and development. We found that different subpopulations of ETV2+ progenitors showed distinct differentiation potentials toward cardiomyocytes and ECs. The segregation of endothelial and myocardial fates was controlled in an *ETV2*-threshold dependent manner. scRNAseq analysis described the transcriptional dynamics during cardiomyocytes and ECs development and maturation, which correlated with the time-course of bulk RNAseq analysis. Early and late progenitors for both lineages were also identified. In summary, this study depicted the developmental path and segregation of human cardiomyocytes and ECs from CM in hiPSC.

RESULTS

Single-cell profiling of co-differentiation of cardiomyocytes and endothelial cells from hiPSCs

To investigate the developmental path of ECs and cardiomyocytes from hiPSC (Figure 1A), we carried out scRNAseq) on day 6 of differentiation. A total 5107 and 3743 cells were analysed from two independent differentiations (Figure S1A-B) and visualized using Uniform Manifold Approximation and Projection (UMAP). Unsupervised cell clustering resulted in 5 clusters (Figure S1C). Cells in cluster 4 showed high expression of pluripotent markers, such as *POU5F1* and *NANOG* (Figure S1D) and were therefore excluded from downstream analysis (Figure S1E). Cells from two independent differentiations clustered similarly, although replicate 2 had fewer cells in cluster 1 (Figure S1F). Cells in clusters 3 and 5 mainly differed in the expression of cell-cycle- and proliferation-related genes (data not shown), with the highest G2M- and Proliferation scores found in cluster 5 compared to cluster 3 (Figure S1G-H). We therefore further reduced the number of clusters to three (C1, C2, C3) by combining clusters 3 and 5 (Figure 1B).

Next, marker genes of each cell cluster were identified in order to determine their identities (Table S1). Cluster C1 was characterized by increased expression of CM and cardiac progenitor genes, such as *MESP1, ISL1, DKK1, SMARCD3, ABLIM1, TMEM88* as well as cell cycle genes *CDK6* and *NEK2* (C1: CM (Figure 1C, F). Cluster C2 was characterized by increased expression of early cardiomyocyte-specific genes, including *MYL4, TNNI1, MYL7, ACTA2, TNNT2, HAND2* and *NKX2-5* (C2: Cardiac progenitors (CPs))(Figure 1D, G). Cluster C3 was characterized by increased expression of EC-specific

genes, such as *CDH5, CD34, KDR, HEY2, TEK, TIE1, ACVRL1, SOX17, ENG, ICAM2, PECAM1* (Endothelial progenitors) (EPs)) (Figure 1E, H). Notably, the *ETV2* transcription factor, a master regulator of hematoendothelial specification, was highly expressed in the EP cell cluster and a small fraction of cells in the CM and CP cell clusters (Figure 1I).

Pseudotime analysis of co-differentiation of endothelial cells and cardiomyocytes from hiPSCs

Diffusion pseudotime analysis (Haghverdi et al., 2016) at day 6 of differentiation showed segregation into three developmental branches that corresponded to CM, CP and EP cell clusters (Figure 2A). Developmental trajectories of CPs and EPs originating at the CM branch. Based on the pseudotime analysis, the EP branch progressed further in development compared to the CP branch. *MESP1* was highly expressed in the CM branch and at an early time-point of the EP and CP branches (Figure 2B). Cardiac makers *MYL4, TNNI1* and EC markers *CDH5, PECAM1* were expressed specifically in the CP and EP branches respectively (Figure 2B). Genes related to cardiac or muscle functionalities, like *ACTC1, GATA4, PKP2, HAND1* and *PDLIM5* were expressed at the end of CM branch and further increased in the CP branch (Figure S2A). Endothelial specific genes *CDH5, PECAM1, CD34, KDR* and *SOX17* increased in pseudotime from the CM branch to the EP branch (Figure S2B). *ETV2* was highly expressed in the early EP branch and in a small number of cells across CPs and the late CM branch (Figure 2B). *ETV2* showed transient expression around 2.19 to 2.49 pseudotime time label with peak expression around 2.25 pseudotime time label (Figure 2C, S2A-B). Analysis of *ETV2* expression in the period of 2.19 to 2.49 pseudotime time label showed highest *ETV2* expression and percentage of *ETV2+* cells in the EP branch (Figure 2D and Figure S2C). *ETV2* expression was comparable in *ETV2+* cells in all three developmental branches (Figure S2D). Expression of *ETV2* downstream target genes, such as *TAL1*, *GATA2* and *LMO2* (Liu et al., 2015) was induced after *ETV2* and correlated with the increased *ETV2* expression (Figure 2E, F). Analysis of DEGs in *ETV2+* cells in CP and EP developmental branches showed enrichment in cardiomyocyte and endothelial cell specific genes indicating that the cells acquired these respective identities (Figure 2G, Table S2) and have further segregated from *ETV2+* cells in the CM developmental branch (Figure S2E-F).

Generation and characterization of genetically encoded ETV2^{mCherry} fluorescent hiPSC reporter line

In order to follow *ETV2* expression during co-differentiation of endothelial cell- and cardiomyocytes differentiation of hiPSCs in real-time, a P2A-mCherry fluorescence reporter with nuclear localization signal (NLS) and neomycin selection cassette was inserted into the endogenous *ETV2* locus before the stop codon using CRISPR/Cas9-facilitated homologous recombination (Figure S3A). Targeted hiPSC clones after

Neomycin selection and excision of the selection cassette were confirmed by PCR (Figure S3A, B) and Sanger sequencing. Clones with ETV2^{mCherry} targeted into both alleles were selected for downstream analysis (Figure S3C). Time-lapse imaging of differentiating cultures showed the appearance of nuclear mCherry from day 4 of differentiation followed by the appearance of endothelial cell markers, such as VE-cadherin (CD144) and lectin (Figure S3D-E and Video S1). mCherry protein remained for a longer period than the endogenous *ETV2* mRNA expression (Figure S3F-G).

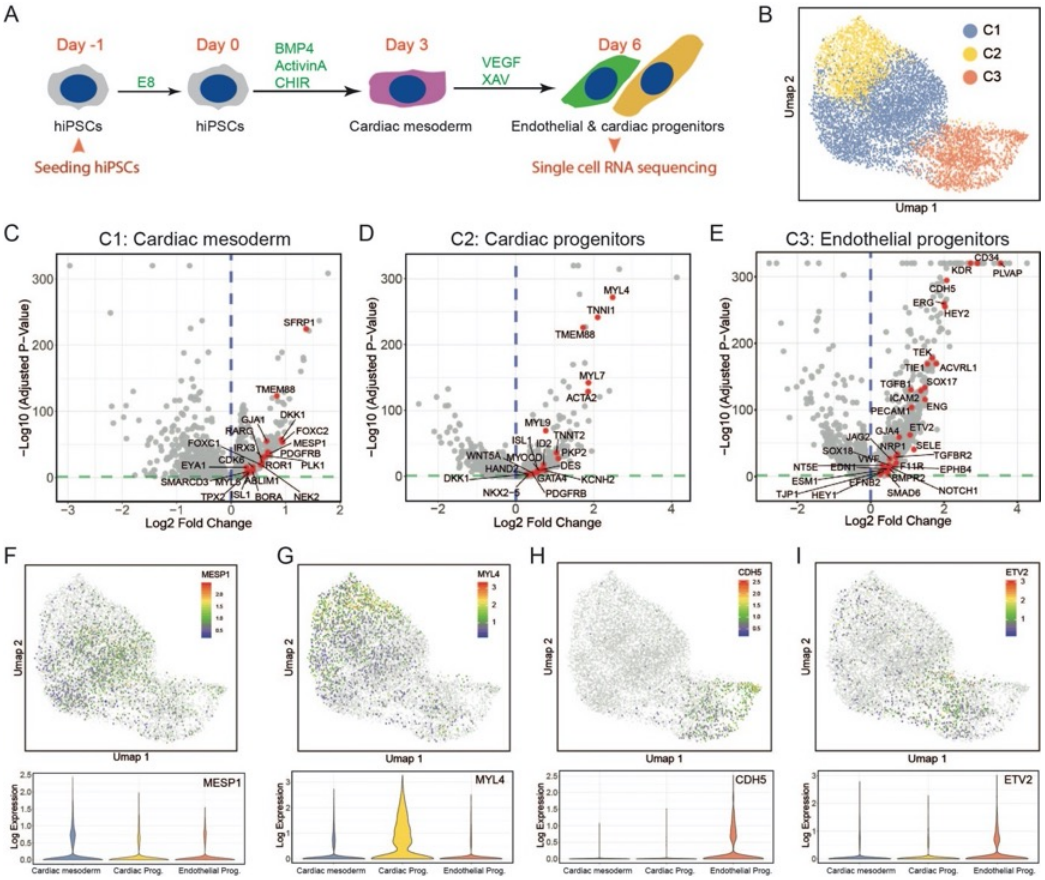

Figure 1. Single-cell RNA sequencing analysis of endothelial and cardiac lineages during co-differentiation of hiPSCs (A) Schematic overview of CMEC differentiation protocol from day -1 to day 6. Cells were collected for scRNAseq on day 6. (B) scRNAseq data of day 6 visualized using UMAP. Three clusters (C1-C3) of cells were identified. (C-E) Volcano plots show fold change and p values of all genes in each cluster compared to background (all the rest of cells). Representative significant up-regulated genes ($P_{adjusted}$<0.05 & fold change>1.2) are labelled in red. (F-G) *MESP1*, *MYL4*, *CDH5* and *ETV2* expression (log_2 transformed) in individual cells (UMAPs) and three clusters (violin plots).

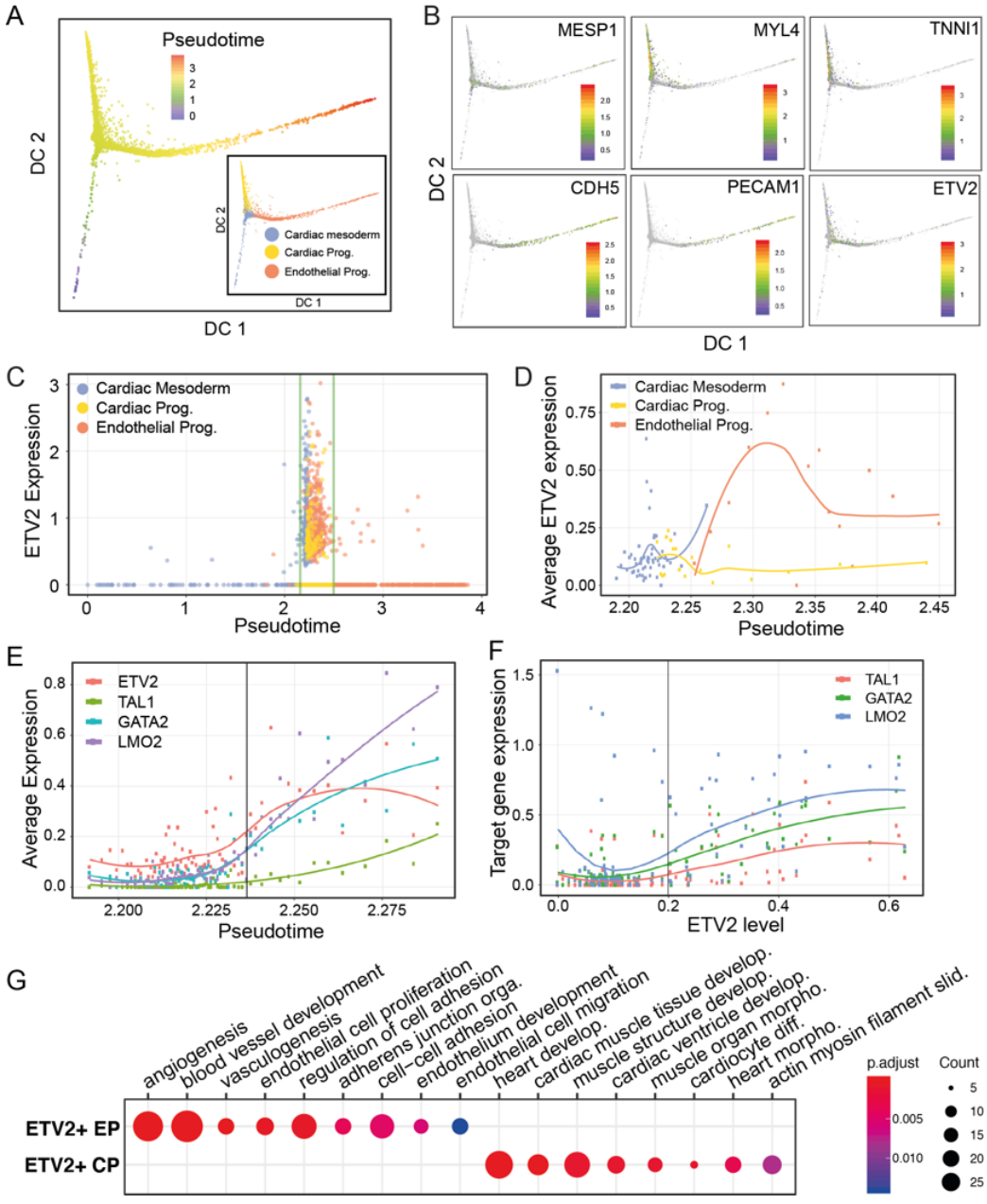

Figure 2. Pseudotime analysis for EC and cardiomyocytes co-differentiation from hiPSCs (A) Dimensionality reduction of scRNAseq data of day 6 using diffusion map. Cell identities are labeled with different colors (bottom right) based on clusters identified in Figure 1B. (B) Diffusion maps show expression patterns of *MESP1, MYL4, TNNI1, CDH5, PECAM1* and *ETV2*. (C) Expression (log_2 transformed) of *ETV2* following pseudotime in CM, CPs and EPs branches. Pseudotime 2.19 and 2.49 were labeled with green lines to show main *ETV2* expression time window. (D) Average expression of *ETV2* of each branch between pseudotime 2.19 to 2.49. Bins

of pseudo time with equal number of cells (50 cells) were constructed and the average expression was calculated among these cells. (E) Average expression of *ETV2* and its direct target genes *TAL1*, *GATA2*, *LMO* in CM and EPs ceslls between pseudotime 2.19 to 2.49. Bins of pseudo time with equal number of cells (50 cells) were constructed and the average expression was calculated among these cells. (F) Correlation of ETV2 level with ETV2 target genes' expression in CM and EPs cells between pseudotime 2.19 to 2.49. Bins of pseudo time with equal number sof cells (50) were constructed and the average expression was calculated among these cells. (G) GO enrichment analysis of differentially expressed genes (DEGs) between ETV2+ EPs and ETV2+ CPs. 128 and 136 genes were upregulated in ETV2+ EPs and CPs respectively ($P_{adjust} < 0.05$). A complete list of GOs can be found in Table S2. Color represents the $P_{adjusted}$ of enrichment analysis and dot size represents count of genes mapped to the GO.

ETV2^{mCherry} fluorescent reporter line recapitulates developmental trajectories during co-differentiation of endothelial cells and cardiomyocytes from hiPSCs

Bulk RNA sequencing was next performed on ETV2^{mCherry}+CD144+ (DP) and ETV2^{mCherry}+CD144- (SP) cells sorted on day 4, 5, 6 and 8 of endothelial cell- and cardiomyocyte co-differentiation from at least 3 independent differentiations (Figure 3A). Principal component (PC) analysis showed separate clustering of DP and SP cells with progressive change over the time (Figure 3B). Mapping of bulk RNAseq samples on scRNAseq diffusion pseudotime plot resulted in co-aligning of DP cells on the EP branch and SP cells on the CP branch (Figure 3C). Notably day 8 SP cells clustered further away from the differentiating cells in the CP branch suggesting a more advanced differentiation state (Figure 3C).

Consensus clustering of the most variable genes across DP or SP cells (3000 genes for each cell type) resulted in three gene clusters with distinct gene expression profiles D1-D3 (DP cells) and S1-S3 (SP cells) (Figure S4A-B, Table S3). Cluster D1 (1226 genes) was characterized by increased expression of of respective genes over time, cluster D2 (1127 genes) was characterized by decreased expression of of respective genes at day 5 of differentiation and cluster D3 (647 genes) was characterized by the decreased expression of of respective genes on day 8 of differentiation (Figure 3D). Gene Ontology (GO) analysis of cluster D1 showed enrichment in terms for angiogenesis, Notch signalling pathway, transforming growth factor beta receptor signalling pathway, receptor-mediated endocytosis and developmental maturation (Figure 4F, Table S4). Specifically, angiogenesis-related genes (*CDH5*, *TIE1*, *TEK*, *EFNB2*, *SOX18*, *VEGFB*, *LEPR)*, Notch and transforming growth factor beta receptor signalling pathway related genes (*COL1A2*, *NOTCH1*, *HES4*, *DLL4*, *JAG2*, *HEY1*, *NOTCH3*, *NOTCH4*, *TGFBR2*, *EGF)* and heart valve morphogenesis related genes (*SMAD6*, *EFNA1*, *GATA5*, *HEY2*, *EMILIN1*, *NOS3*, *GATA3*) were upregulated over the course of differentiation in DP cells (Figure 3H, S4C).

Cell cycle related genes were enriched in cluster D2 (*ITGB1*, *CDK4*, *CCND1*, *CDK2AP2*, *MYC*, *CDC6*) (Figure S4D). Cell proliferation related genes and fatty-acyl-CoA biosynthetic process genes were enriched in cluster D3 (*ACLY*, *FASN*, *ELOVL1*, *SLC25A1, ACSL3, ACSL4*) (Figure S4E).

Cluster S1 (936 genes) was characterized by the increased expression of respective genes over the time, cluster S2 (746 genes) was characterized by the decreased expression of respective genes at day 8 of differentiation and cluster S3 (1318 genes) was characterized by the decreased expression of cell type specific genes later on day 5 of differentiation (Figure 3E). GO analysis of genes in cluster S1 showed enrichment in terms for related to the heart development and function (Figure 3G, Table S4). Specifically, cardiac chamber and cardiac muscle development related genes were upregulated over the course of differentiation in SP cells (*MYH6, HAND1, MYH10, TNNT2, NKX2-5, ISL1, TNNC1, MYOD, LMO4* and *HEY1, MYL7, MYL4, ACTA2*, *KCNH2*) (Figure 3I, S4F). Mitotic nuclear division genes (TPX2, CDC20, NEK2, PLK1, PRC1 and CDC25C) were enriched in cluster S2 (Figure S4F). Transcription and translation process related genes (*SF1, SNRPE, DDX23, RRP1B* and *PRMT5*) were enriched in cluster S3 (Figure S4G).

Next, mean gene expression per cell for each gene cluster (D1-D3 and S1-S3) from the bulk RNA sequencing were calculated and mapped on scRNAseq UMAP and diffusion pseudotime plots (Figure 3J-M, S4H-K). Genes in cluster D1 were specifically expressed in the EP cluster (Figure 3J) and showed increased expression along the pseudotime EP branch (Figure 3L). Genes in cluster S1 were highly expressed in the CM and CP clusters (Figure 3K) and showed increased expression along the pseudotime CP branch (Figure 3M). Genes in clusters D2-D3 and S2-S3 showed similar patterns of expression with the highest expression in the CM cluster (Figure S4H, I) and early pseudotime EP branch (Figure S4J-K).

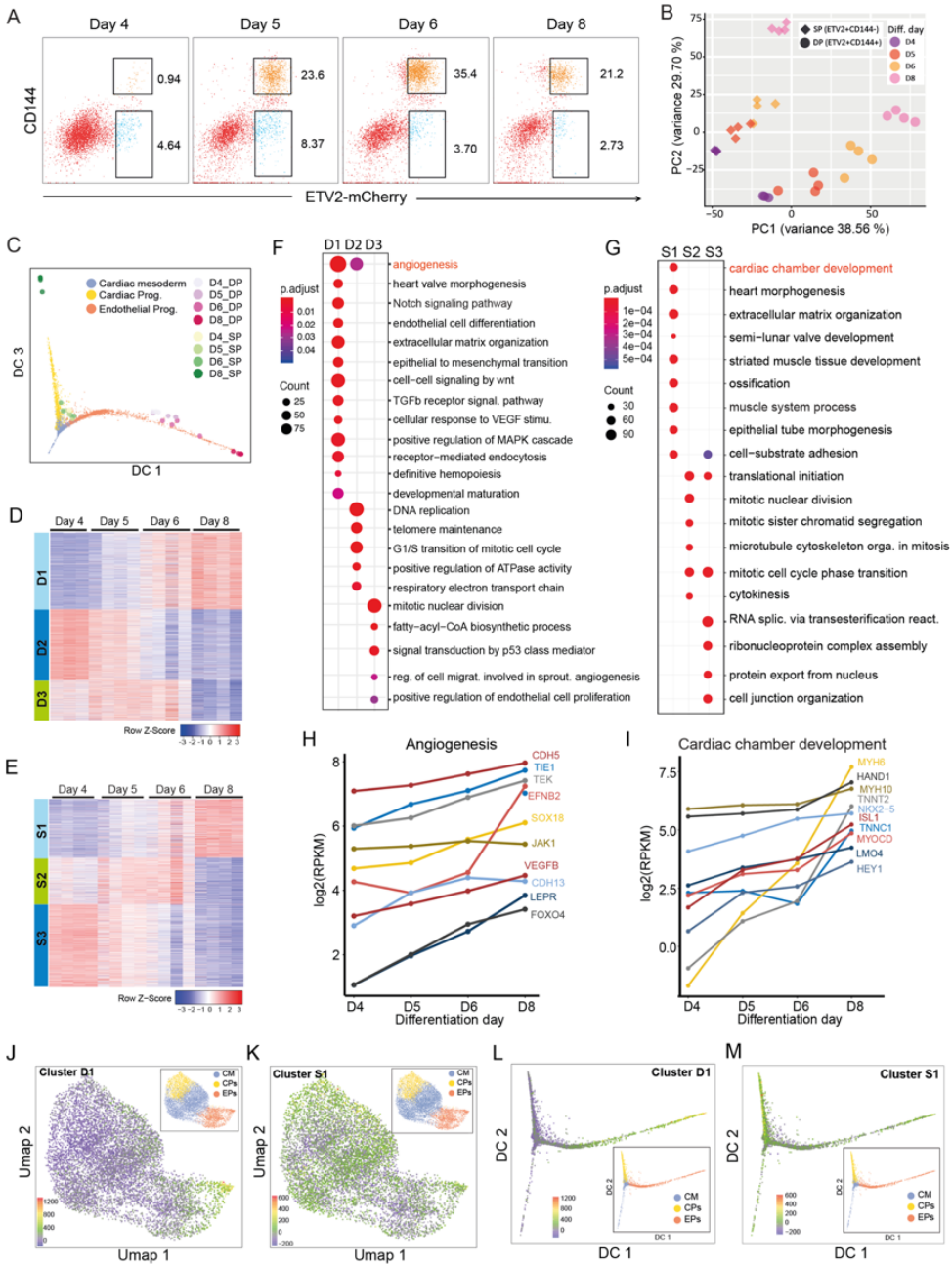

Figure 3. Transcriptomic analysis of EC and cardiac cell development from hiPSCs using a hiPSC ETV2^{mCherry} reporter line (A) FACS analysis of CD144 and ETV2^{mCherry} expression at day 4, 5, 6 and 8 of differentiation. ETV2+CD144+ (DP) and ETV2+CD144- (SP) cells were sorted on each day for bulk RNA sequencing. (B) PCA of all sorted DP and SP samples collected from three or four independent differentiations. (C) Diffusion maps of scRNAseq analysis and its overlay with all bulk RNAseq samples collected on day 4, 5, 6 and 8. Different clusters of cells or bulk samples are labelled with different colours. (D-E) Gene expression pattern in all DP (D) and SP (E) cells.

3000 most variable genes across all DP or SP samples were identified and grouped into three clusters by Consensus clustering. Genes in each cluster can be found in Table S3. Color scale represents relative expression (row z-score). (F-G) GO enrichment analysis of each gene cluster of DP (C) and SP (D) samples. Representative GOs are shown, while complete list of GO can be found in Table S4. Color represents the $P_{adjusted}$ of enrichment analysis and dot size represents count of genes mapped to the GO. (H-I) Representative genes mapped to representative GOs of cluster D1 (E) and S1 (F) and their expression levels from day 4 to day 8 are shown. (J-K) Standardized sum of expression of genes in cluster D1 (F) and S1 (G) was calculated for each cell based on scRNAseq data and visualized in UMAPs. Color represents standardized sum of expression value. (L-M) Standardized sum of expression of genes in cluster D1 (I) and S1 (J) was calculated for each cell based on scRNAseq data and displayed in diffusion maps. Color represents standardized sum of expression value.

Diversification of endothelial and cardiomyocyte cell lineages during simultaneous co-differentiation from hiPSCs

To validate ETV2^{mCherry} reporter, ETV2^{mCherry}+ and ETV2^{mCherry}- cells were sorted on day 5 of differentiation and differentiated toward ECs in the presence of VEGF (Figure 4A-B). FACs analysis on day 5 after sort resulted in pure population of CD144+CD31+ ECs (>90%) from ETV2^{mCherry}+ cells with a small fraction of ECs (10-15%) developed from ETV2^{mCherry}- cells (Figure 4C-D). Cells derived from ETV2^{mCherry}+ cells expressed endothelial-specific markers, as observed by qPCR and immunofluorescence analysis (Figure 4E-H). Functionally, ECs differentiated from ETV2^{mCherry}+ cells upregulated pro-inflammatory markers, such as ICAM-1 and E-Selectin upon TNF-α stimulation (Figure 4I-L), as shown previously for hiPSC-ECs (Halaidych et al, 2018, SCR).

ETV2^{mCherry} DP, SP and double negative (DN) cells were next sorted at day 7 of differentiation and further cultured in the presence of VEGF until day 18 (Figure 4A-M). DP cells differentiated into CD144+CD31+ ECs (>80%) (Figure 4N-O). In contrast, SP and DN cells differentiated into cTnT+ cardiomyocytes (>50%) with very few ECs being detected (Figure 4N-O). Interestingly, cardiomyocytes derived from SP cells seem to proliferate more and formed a monolayer composed of a contracting cell sheet, while cardiomyocytes from DN cells proliferated to a less extend and with only a number of separated cell clusters contracting (Video S2). Almost all DP cells on day 18 expressed EC marker CD31, while only few numbers of cells derived from SP and DP cells were positive for CD31 (Figure 4P-R). Most cells derived from SP and DN expressed cardiomyocyte-specific α-Actinin and cTnT and showed typical sarcomeric structures (Figure 4Q-R). A small number of SP and DN-derived cells were also positive for the smooth muscle cell marker SM22 while negative for cardiac markers (data not shown).

Furthermore, SP cell fraction-derived α-Actinin positive cardiomyocytes were positive for SM22 indicating possibly their immaturity (Figure 4R).

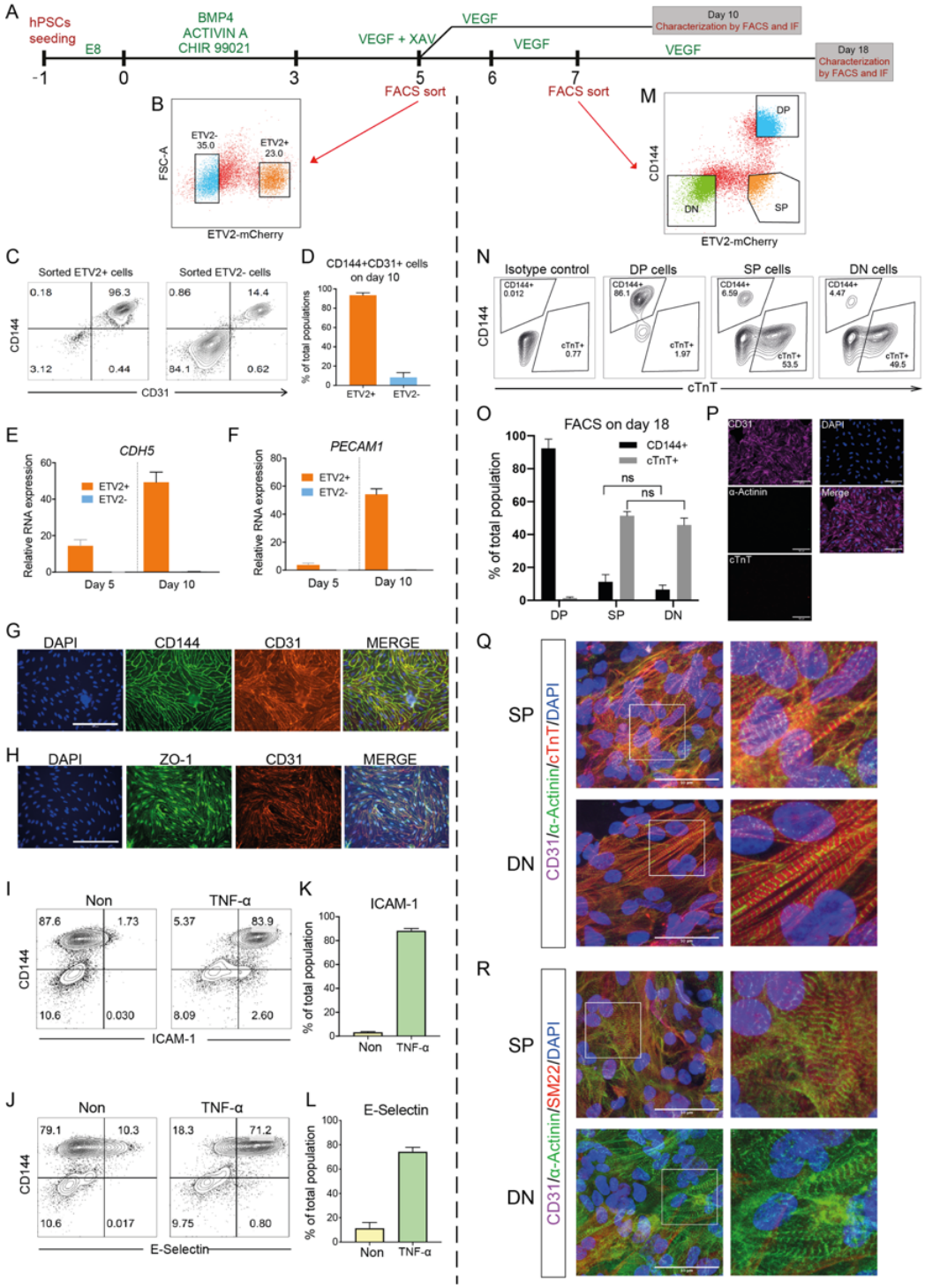

Figure 4. Diversification of endothelial and cardiomyocyte cell lineages during simultaneous

co-differentiation from hiPSCs (A) Schematic of CMEC differentiation protocol and FACS sort. ETV+ and ETV2- cells were sorted on day 5 and cultured in VEGF until day 10. DP, SP, DN were sorted on day 7 and cultured in VEGF until day 18. (B) Representative FACS analysis ETV2-mCherry on day 5 and gates for FACS sort of ETV2+ and ETV2- cells are shown. (C) FACS analysis of endothelial markers CD144 and CD31 on day 10 of sorted ETV2+ and ETV2- cells. (D) Quantification of the percentage of CD144+CD31+ cells in total populations on day 10 of sorted ETV2+ and ETV2- cells. (E-F) Quantification of gene expression of *CDH5* and *PECAM1* in sorted ETV2+ and ETV2- cells on day 5 and day 10. (G-H) Immuno-staining of CD144, CD31 and cell-cell junctional marker ZO-1 on day 10. Scale bar 200 µm. (I-J) FACS analysis of ICAM1, E-Selectin and CD144 for sorted ETV2+ cells on day 10. Cells were stimulated with TNF-α for 24 h before FACS. (K-L) Quantification of percentage of CD144+ICAM-1+ (K) and CD144+E-Selectin+ (L) cells in total population on day 10. (M) FACS analysis of CD144 and ETV2-mCherry expression on day 7. DP, SP and DN cells were gated and sorted. (N) FACS analysis of CD144 and cardiomyocytes marker cTnT on day 18 of sorted DP, SP and DN cells. Isotype antibodies were included as control. (O) Quantification of CD144+ ECs and cTnT+ cardiomyocytes on day 18 of DP, SP and DN cells. (E) Immuno-staining of CD31, a-Actinin, cTnT and DAPI on day 18 of DP cells. Scale bar 50 µm. (F) Immuno-staining of CD31, a-Actinin, cTnT, SM22 and DAPI on day 18 of SP and DN cells. Scale bar 50 µm. Error bars are ± SD of three independent experiments in (D-F, K-L, O).

DISCUSSION

In this study, we characterized the development of ECs and cardiomyocytes during their co-differentiation from hiPSC (Giacomelli et al., 2017) using their bulk- and scRNAseq profiles. Our major question was whether the ECs and cardiomyocytes actually had a common precursor as assumed previously on the basis of studies in mice or whether they differentiated in parallel.

Analysis showd that all cells could be divided into three principal clusters on day 6 of differentiation, corresponding to mesodermal, early myocardial and endothelial identities. Myocardial and endothelial lineages had a common origin in *MESP1*+ mesoderm cells in this *in vitro* system, in agreement with previous *in vivo* studies (Lescroart et al., 2018; 2014). Genes involved in early cardiomyocytes specification, including *SMARCD3* (Devine et al., 2014), *TMEM88* (Palpant et al., 2013) and *ISL1* (Quaranta et al., 2018) were highly expressed in CMs. Specifically, *TMEM88* was shown previously to promote cardiomyocyte but inhibit EC differentiation from multipotent cardiac mesoderm (Palpant et al., 2013), which was indeed absent in the endothelial lineage here. EPs differentiated from hiPSCs showed a clear EC identity and were fully functional based on their inflammatory response upon TNFα stimulation and endocytosis of DiI-AcLDL. Notably, they also expressed a number of cardiac markers like *MEOX2*, *GATA4*, *GATA6* and *ISL1,* suggesting an cardiac specific EC identity, which

corelated with our previous findings (Giacomelli et al., 2017). Myocardial cells, on the other hand, were still early progenitors on day 6 of the differentiation as no functionally contracting cardiomyocytes were observed yet at this stage.

ETV2 was identified as a specific marker of EP cluster in our scRNAseq dataset, corresponding to its fundamental role in hemagiogenic development (Garry et al., 2017). Interestingly, *ETV2* expression was also observed in a small population of CM and CP cell clusters. Moreover, the *ETV2*+ population on day 6 of differentiation were highly heterogenous, being composed of mesoderm, cardiac and endothelial lineages. Pseudotime analysis suggested that *ETV2*+ cardiomyocytes and endothelial cells both originated from *ETV2*+ CM. ECs had higher percentage of *ETV2*+ cells as well as average *ETV2* expression than the other two lineages. Although average *ETV2* levels in the *ETV2*+ subpopulation from different lineages were comparable over the whole time period, a short pulse of *ETV2* upregulation was specifically observed in endothelial *ETV2*+ cells during the differentiation (Figure S2D). In addition, expression of direct target genes of *ETV2* were triggered once an *ETV2* expression threshold was reached. This was also observed in previous reports during hemagiogenic specification (Zhao and Choi, 2019; 2017). These results suggested an *ETV2* pulse- and threshold dependent segregation of cardiomyocytes and ECs from their common mesoderm precursors.

An ETV2^{mCherry} hiPSC reporter line generated to track, isolate and characterize *ETV2*+ cells during CMEC differentiation showed that ETV2^{mCherry}+ cells could give rise to both Lectin+ ECs and Lectin- non-ECs on day 8 (Figure 2C), suggesting that not all ETV2 expressing cells could give rise ECs. Through the bulk RNAseq analysis for ETV2^{mCherry}+ ECs (DP) and non-ECs (SP) on different days of the differentiation, DP and SP cells were found to acquire more specific endothelial and myocardial identities as well as downregulate cell cycle-related genes from day 4 to day 8, indicating further development and maturation of these cells. Notably, although *CDH5* was expressed as early as day 4, several key angiogenesis and Notch signaling pathway genes, like *LEPR*, *FOXO4*, *DLL4*, *NOTCH4* and *EGF*, only started to be expressed after day 4 indicating a specified EC fate but immature progenitor identity on day 4. These relatively late expressed genes could potentially be used as markers to distinguish early and late ECs during development *in vitro* or *in vivo*. Genes involved in heart development and definitive hematopoiesis were also upregulated during ECs development (Figure 3B), suggesting cardiac endothelial- and probably a mixture of hemogenic endothelial identity of these ECs, while their hematopoietic potential should be further studied although this is beyond the scope of this study. In terms of SP cells, cardiac and muscle development related genes were highly enriched and upregulated during its development, suggesting a myocardial fate of these cells. Although SP cells had already committed to a cardiac fate by upregulation of cardiac genes *HAND1*, *MYH10*, *NKX2-5*, *ISL1*, *TNNC1*, *MYOCD* and *LMO4*, some crucial genes for cardiomyocytes were still

absent on day 4, including *MYH6* and *TNNT2*. *MYH6* encodes the major cardiomyocyte thick filament protein MHC-α and *TNNT2* was routinely used as a cardiomyocyte marker. Both genes are essential for cardiomyocyte contractility and started to be expressed after day 4. Their relatively late expression could allow us to identify early and late cardiac progenitors during cardiac development in future studies.

scRNseq data on day 6 showed that endothelial and cardiac specific genes were upregulated along the EP- and CP development routes using unbiased pseudotime analysis, which correlated with the bulk RNAseq analysis and confirmed the specification and maturation process of these two lineages. Most importantly, DP and SP samples from bulk RNAseq were overlaid well with the EP and CP branches according to their differentiation times within the diffusion plot of pseudotime analysis. This further confirmed the endothelial and myocardial fates of DP and SP cells, respectively. However, it is worth mentioning that DP and SP did not fully correspond to EP and CP clusters. The large number of *CDH5* negative cells identified in the early EP branch was not present in sorted DP population. Therefore, all DP samples were mapped to the latter half of the EP branch (Figure 3F). Although all SP samples were mapped to the CP branch due to their common myocardial identities, day 8 SP cells were developed further beyond the end of CPs branch. This suggested that cardiac cells on day 8 were more mature and could not be captured by scRNseq on day 6, which correlated well with the dramatic increase of cardiac genes from day 6 to day 8 in SP samples (Figure 4B).

Interestingly, we found there were two subpopulations of ETV2+ progenitors with different differentiation potentials. DP cells sorted on day 7 gave rise to endothelial cells only in the presence of VEGF, as expected. However, the majority of SP cells, also sorted on day 7, consistently gave rise to contracting cardiomyocytes under the same culture conditions. This provided direct evidence that the SP cells were cardiomyocyte progenitors. More importantly, it demonstrated that both ECs and cardiomyocytes could be derived from ETV2+ progenitors, confirming the presence of a common precursor implied by our earlier studies (Giacomelli et al, 2017). Notably, though, only a few (~5%) ECs could be derived from SP and DN cells, which could be derived from multipotent progenitors with delayed differentiation. Although SP and DN cells had similar differentiation potentials with similar efficiencies, the SP population had a higher cell growth after sorting (data not shown). Cardiomyocytes derived from SP cells also showed higher contractility compared to DN-derived cells (Video S2). This difference could be due to either the different cell densities (growth rates) or their different development origins (FHF vs. SHF). More work is needed to establish the identity of cardiomyocytes from SP and DN populations definitively.

In summary, this study used time-course bulk RNAseq and scRNAseq analysis to investigate the co-differentiation of cardiomyocytes and ECs from hiPSCs. scRNAseq

revealed transcriptional dynamics during early myocardial and endothelial lineage segregation from their common *MESP1*+ CM progenitors. We found that both cardiomyocytes and ECs could be specified from ETV2+ precursors in an *ETV2*-threshhold-depedent manner, which may reflect differentiation *in vivo* process from multipotent CPCs. Overall, this study provides insights into the developmental of different human cardiac lineages and a rich data set for comparison with lineage reporters in mice.

MATERIAL AND METHODS

hiPSC culture

The NCRM1 hiPSC line (NIH) was used in this study. This hiPSC control line was cultured in TeSR-E8 on Vitronectin XF and was routinely passaged once a week using Gentle Cell Dissociation Reagent (all from Stem Cell Technologies). Prior to targeting NCRM1 hiPSCs were passaged as a bulk on feeders in hESC-food medium. RevitaCell (Life Technologies) was added to the medium (1:200) after every passage to enhance viability after single cell passaging with TrypLE (Life technologies).

Generation of hiPSC reporter line using CRISPR/Cas9

The p15a-cm-hETV2-P2A-NLS-mCherry-neo repair template plasmid was generated using overlap PCR and restriction-based cloning and ligation. The ETV2 homology arms were amplified from genomic DNA and the neomycin cassette flanked by FRT sites was amplified from the p15A-cm-PAX3-venus-neo-1kb plasmid (from Konstantinos Anastassiadis). P2A-NLS-mCherry double-stranded DNA fragment was ordered from IDT. The sgRNA/Cas9 plasmid was modified from SpCas9-2A-Puro V2.0 plasmid (Addgene, Feng Zang).

NCRM1 hiPSCs were passaged with ratio 1:2 or 1:3 into 60 mm dishes to reach 60-70% confluency the next day for transfection. 20 µl lipofectamine (Invitrogen), 8 µg of repair template and 8 µg of sgRNA/Cas9 plasmid were diluted in 600 µl of Opti-MEM and added to each 60 mm dish. After 18 hours the medium was changed to hESC-food. After another 6 hours G-418 (50 µg/ml) selection was started and was kept for 1 week. Survived cells were cultured in hESC-food and passage into 6-well plate for the transfection of Flp recombinase expression vector to remove the neomycin cassette. 300 µl of Opti-MEM containing 10 µl lipofectamine and 4 µg CAGGs-Flpo-IRES-puro plasmid was added per well for 18 hours. Puromycin (0,5 µg/ml) selection was started 24 hours post transfection and lasted for 2 days. Once recovered, cells were passage into 96-well format for clonal expansion via limited dilution. Targeted clones were identified by PCR and sequencing. Primers outside the ETV2 homology arms and primers inside the targeting construct were used to confirm on-target integration. The absence of mutation within inserted sequence and untargeted allele were confirmed by Sanger

sequencing (BaseClear).

Endothelial and myocardial lineages co-differentiation from hiPSCs

Endothelial and cardiac cells were induced from hiPSCs in a monolayer using CMEC protocol as described previously (Giacomelli et al., 2017). Briefly hiPSCs were split with a 1:12 ratio and seeded on 6-well plates coated with 75 µg/mL (growth factor reduced) Matrigel (Corning) on day -1. At day 0, cardiac mesoderm was induced by changing TeSR-E8 to BPEL medium (Ng et al., 2008), supplemented with 20 ng/mL BMP4 (R&D Systems), 20 ng/mL ACTIVIN A (Miltenyi Biotec) and 1.5 µM CHIR99021 (Axon Medchem). At day 3, cells were refreshed with BPEL supplemented 5 µM XAV939 (Tocris Bioscience) and 50 ng/ml VEGF (R&D Systems). From day 6 onwards, cells were refreshed every 3 days with BPEL medium supplemented with 50 ng/ml VEGF.

Fluorescence-activated cell sorting

For FACS sorting on day 5 of CMEC protocol, ETV2-mCherry positive and negative cells were sorted using FACSAria III (BD-Biosciences). Around 20k cells/cm2 were seeded on Fibronectin (from bovine plasma, 5µg/ml, Sigma Aldrich) coated plates. Cells were cultured in BPEL supplemented with VEGF (50 ng/ml) until day 10. The medium was refreshed every 3 days.

For FACS sorting on day 7 of CMEC protocol, VEC+mCherry+ (DP), VEC-mCherry+ (SP) and VEC-mCherry- (DN) cells were sorted using FACSAria III. 1 million cells were seeded in each well of Matirgel-coated 12-well plate in BPEL supplemented with VEGF (50 ng/ml) and RevitaCell (1:200). Cells were refreshed 24 h after seeding and every three days afterwards with BPEL supplemented with VEGF (50 ng/ml).

Immunofluorescence staining and imaging

Cultured cells were fixed in 4% paraformaldehyde for 15 min, permeabilized for 10 min with PBS containing 0.1% Triton-X 100 (Sigma-Aldrich) and blocked for 1h with PBS containing 5% BSA (Sigma-Aldrich). Then cells were stained with primary antibody overnight at 4oC. The next day cells were washed three times (20 min each time) with PBS. After that cells were incubated with fluorochrome-conjugated secondary antibodies for 1h at room temperature and washed three times (20 min each time) with PBS. Then cells were stained with DAPI (Life Technologies) for 10 min at room temperature and washed once with PBS for 10min. Both primary and secondary antibodies were diluted in 5% BSA/PBS. Images were taken with EVOS FL AUTO2 imaging system (ThermoFischer Scientific) with 20x objective. For staining of ECs and cardiomyocytes on day 18, images were taken with a Leica SP8WLL confocal laser-scanning microscope using a 63x magnification objective and Z-stack acquisition. Details of all antibodies that were used are provided in Table S5.

AcLDL uptake assay

Alexa FluorTM 594 AcLDL (ThermoFisher Scientific) was diluted in lipid-free BPEL to a final concentration of 5 μg/ml and added to ECs on differentiation day 10. After 4 hours incubation in 37 °C, wash cells once with lipid-free BPEL medium. Stain nucleus with NucBlue™ Live ReadyProbes™ for 20 min in 37 °C and then take images.

FACS analysis

Cells were washed once with FACS buffer (PBS containing 0.5% BSA and 2 mM EDTA) and stained with FACS antibodies for 30 min at 4oC. Samples were washed once with FACS buffer and analyzed on MACSQuant VYB (Miltenyi Biotech) equipped with a violet (405 nm), blue (488 nm) and yellow (561 nm) laser. The results were analyzed using Flowjo v10 (FlowJo, LLC). Details of all fluorochrome conjugated FACS antibodies are provided in Table S5.

Quantitative Real-Time Polymerase Chain Reaction (qPCR)

Total RNA was extracted using the NucleoSpin® RNA kit (Macherey-Nagel) according to the manufacturer's protocol. cDNA was synthesized using an iScript-cDNA Synthesis kit (Bio-Rad). iTaq Universal SYBR Green Supermixes (Bio-Rad) and Bio-Rad CFX384 real-time system were used for the PCR reaction and detection. Relative gene expression was determined according to the standard ΔCT calculation and normalized to housekeeping genes (mean of HARP and RPL37A). Details of all primers used are provided in Table S6.

Bulk RNA sequencing and analysis

Cells were sorted on differentiation day 4, 5, 6 and 8 for bulk RNA-Seq. Total RNA was extracted using the NucleoSpin® RNA kit (Macherey-Nagel). Whole transcriptome data were generated at BGI (Shenzhen, China) using the Illumina Hiseq4000 (100bp paired end reads). Raw data was processed using the LUMC BIOPET Gentrap pipeline (https://github.com/biopet/biopet), which comprises FASTQ preprocessing, alignment and read quantification. Sickle (v1.2) was used to trim low-quality read ends (https://github.com/najoshi/sickle). Cutadapt (v1.1) was used for adapter clipping (Martin, 2011), reads were aligned to the human reference genome GRCh38 using GSNAP (gmap-2014-12-23) (Wu and Nacu, 2010) (Wu and Watanabe, 2005) and gene read quantification with htseq-count (v0.6.1p1) against the Ensembl v87 annotation(Yates et al., 2016). Gene length and GC content bias were normalized using the R package cqn (v1.28.1) (Hansen et al., 2012). Genes were excluded if the number of reads was below 5 reads in ≥90% of the samples. The final dataset comprised gene expression levels of 31 samples and 22,419 genes.

Differentially expressed genes were identified using generalized linear models as implemented in edgeR (3.24.3) (Robinson et al., 2010). P-values were adjusted using the Benjamini-Hochberg procedure and FDR ≤ 0.05 was considered significant. Analyses were performed using R (version 3.5.2). PCA plot was generated with the built-in R functions prcomp using transposed normalized RPKM matrix. Correlation among samples was calculated using cor function with spearman method and the correlation heatmap was generated with aheatmap function (NMF package).

Gene clusters were calculated with CancerSubtypes package (Xu et al., 2017). The top 3000 most variable genes across all chosen samples were identified based on the most variant Median Absolute Deviation (MAD) using FSbyMAD function, then z_score normalization was performed for each gene. K clusters were calculated using the K-means clustering of euclidean distance. Clustering was iterated 1000 times for K clusters in the range 2 to 10. Heatmap of genes in all clusters was generated using R basic heatmap function. Gene ontology enrichment for each cluster of genes was performed using compareCluster function of clusterProfiler package (v3.10.1) (Yu et al., 2012) and q ≤ 0.05 was considered significant.

Single-cell RNA sequencing and analysis

Sample preparation and sequencing

Cells were dissociated into single cells on day 6 of CMEC differentiation and loaded into the 10X Chromium Controller for library construction using the Single-Cell 3' Library Kit. Next, indexed cDNA libraries were sequenced on the HiSeq4000 platform. The mean reads per cell were reported as 28,499 in the first replicate and 29,388 in the second replicate.

Pre-processing, clustering and UMAP

Both replicates of day 6 CMEC differentiation were merged into one data set. The average number of detected genes is 2643 and the average total expression per cell is 10382 (Figure S1A-B). Then, undetected genes (> 1 UMI count detected in less than two cells) and cells with low number of transcripts were removed from further analysis (Figure S1A-B). This resulted in 5107 cells for the first replicate, and 3743 cells for the second replicate with 13243 genes each. Expression profiles were normalized with the scran package in R (V 1.10.2) using the method described in (Lun et al., 2016). The 10% most highly variable genes (HVG) for each replicate were calculated with scran after excluding ribosomal genes (obtained from the HGNC website without any filtering for minimum gene expression), stressed genes (van den Brink et al., 2017) and mitochondrial genes. Batch effect correction between replicates was performed with a mutual nearest neighbours (MNN)- based approach (Haghverdi et al., 2018) implemented in the scran package. (Here we used log-transformed and normalized count data with the 10% most HVG that were present in all replicates). The MNN

algorithm was run with d = 30 and k = 20.

Louvain clustering was performed on the shared-nearest-neighbours graph with k = 100 and d = 30, which resulted in 5 clusters. One of the 5 clusters was identified by marker genes as undifferentiated iPSCs and was excluded from further analysis.

Uniform Manifold Approximation and Projection (UMAP) was performed on the principal component scores of the MNN corrected counts using the R package umap (V 0.2.3.1) with 10 nearest neighbours, min_dist = 0.7 and cosine metric.

Cell Cycle and Proliferation Analysis

Cell cycle analysis was performed with the scran package using the cyclone function on normalized counts. For the proliferation analysis a list of known proliferation genes (Whitfield et al., 2006) was used and the average scaled expression per cell was calculated. Based on the cell cycle results two of the four remaining clusters were merged.

Differential expression analysis and identification of cluster maker genes

The R package edgeR (V 3.24.3, 31) (Robinson et al., 2010) was used to perform differential expression analysis. We used raw counts and a negative binomial distribution to fit the generalized linear model. The covariates were comprised of 6 binary dummy variables that indicate the three remaining clusters per replicate and a variable that corresponds to the total number of counts per cell. Finally, p-values for each cluster considering both replicates were obtained and adjusted for multiple hypothesis testing with the Benjamini-Hochberg method.

Pseudo Time Analysis

For the pseudo time analysis, the first 30 principal component scores of the above described MNN corrected normalized counts were used to calculate a diffusion map (R package destiny, V 2.12.0) using default parameters. Then, diffusion pseudo time was calculated with the same package. The root cell was chosen from the cluster identified as cardiac mesoderm via literature markers. In order to display genes in pseudo time the first branch was extracted, which led to two tip cells, one within the endothelial progenitors and the other within cardiac progenitors. In order to display the trend of genes in some plots, bins of pseudo time with equal number of cells were constructed. For each bin the average expression or the average expression percentage was calculated among the cells.

Comparison to bulk RNA-sequencing data

Both replicates of normalized single cell counts were combined with bulk RNA-sequencing data. The MNN approach was used to correct between the two single-cell replicates using the 10% HVG per replicate and the bulk RNA-sequencing data, with d = 30 and k = 20. After batch correction a diffusion map was calculated on the MNN corrected values with default parameters.

Statistics

Statistical analysis was conducted with GraphPad Prism 7 software. Data are represented as mean± SD.

Acknowledgements

S.L. Kloet and E. de Meijer (Leiden Genome Technology Center) for help with 10X Genomics experiments (cell encapsulation, library preparation, single-cell sequencing, primary data mapping, and quality control).

Funding

This project received funding from the European Union's Horizon 2020 Framework Programme (668724); European Research Council (ERCAdG 323182 STEMCARDIOVASC); Netherlands Organ-on-Chip Initiative, an NWO Gravitation project funded by the Ministry of Education, Culture and Science of the government of the Netherlands (024.003.001).

REFERENCES

Buckingham, M., Meilhac, S., Zaffran, S., 2005. Building the mammalian heart from two sources of myocardial cells. Nat. Rev. Genet. 6, 826–835. doi:10.1038/nrg1710

Cai, C.-L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S.L., Chen, J., Evans, S., 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Developmental Cell 5, 877–889. doi:10.1016/s1534-5807(03)00363-0

Caspi, O., Lesman, A., Basevitch, Y., Gepstein, A., Arbel, G., Habib, I.H.M., Gepstein, L., Levenberg, S., 2007. Tissue engineering of vascularized cardiac muscle from human embryonic stem cells. Circ. Res. 100, 263–272. doi:10.1161/01.RES.0000257776.05673.ff

De Val, S., Chi, N.C., Meadows, S.M., Minovitsky, S., Anderson, J.P., Harris, I.S., Ehlers, M.L., Agarwal, P., Visel, A., Xu, S.-M., Pennacchio, L.A., Dubchak, I., Krieg, P.A., Stainier, D.Y.R., Black, B.L., 2008. Combinatorial Regulation of Endothelial Gene Expression by Ets and Forkhead Transcription Factors. Cell 135, 1053–1064. doi:10.1016/j.cell.2008.10.049

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K., Bruneau, B.G., 2014. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. Elife 3, 508. doi:10.7554/eLife.03848

Ema, M., Takahashi, S., Rossant, J., 2006. Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. Blood 107, 111–117. doi:10.1182/blood-2005-05-1970

Ferdous, A., Caprioli, A., Iacovino, M., Martin, C.M., Morris, J., Richardson, J.A., Latif, S., Hammer, R.E., Harvey, R.P., Olson, E.N., Kyba, M., Garry, D.J., 2009. Nkx2–5 transactivates the Ets-related protein 71gene and specifies an endothelial/endocardial fate in the developing embryo. Proc. Natl. Acad. Sci. U.S.A. 106, 814–819. doi:10.1073/pnas.0807583106

Galli, D., Domínguez, J.N., Zaffran, S., Munk, A., Brown, N.A., Buckingham, M.E., 2008. Atrial myocardium derives from the posterior region of the second heart field, which acquires left-right identity as Pitx2c is expressed. Development 135, 1157–1167. doi:10.1242/dev.014563

Garry, D.J., Olson, E.N., 2006. A common progenitor at the heart of development. Cell 127, 1101–1104. doi:10.1016/j.cell.2006.11.031

Giacomelli, E., Bellin, M., Sala, L., van Meer, B.J., Tertoolen, L.G.J., Orlova, V.V., Mummery, C.L., 2017. Three-dimensional cardiac microtissues composed of cardiomyocytes and endothelial cells co-differentiated from human pluripotent stem cells. Development 144, 1008–1017. doi:10.1242/dev.143438

Goldfracht, I., Protze, S., Shiti, A., Setter, N., Gruber, A., Shaheen, N., Nartiss, Y., Keller, G., Gepstein, L., 2020. Generating ring-shaped engineered heart tissues from ventricular and atrial human pluripotent stem cell-derived cardiomyocytes. Nat. communications 11(1), 75. doi:10.1038/s41467-019-13868-x

Haghverdi, L., Lun, A.T.L., Morgan, M.D., Marioni, J.C., 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 36, 421–427. doi:10.1038/nbt.4091

Hansen, K.D., Irizarry, R.A., WU, Z., 2012. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13, 204–216. doi:10.1093/biostatistics/kxr054

Hartogh, den, S.C., Schreurs, C., Monshouwer-Kloots, J.J., Davis, R.P., Elliott, D.A., Mummery, C.L., Passier, R., 2015. Dual reporter MESP1 mCherry/w-NKX2-5 eGFP/w hESCs enable studying early human cardiac differentiation. Stem Cells 33, 56–67. doi:10.1002/stem.1842

Lee, D., Park, C., Lee, H., Lugus, J.J., Kim, S.H., Arentson, E., Chung, Y.S., Gomez, G., Kyba, M., Lin, S., Janknecht, R., Lim, D.-S., Choi, K., 2008. ER71 Acts Downstream of BMP, Notch, and Wnt Signaling in Blood and Vessel Progenitor Specification. Cell Stem Cell 2, 497–507. doi:10.1016/j.stem.2008.03.008

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D., Blanpain, C., 2014. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. Nat Cell Biol 16, 829–840. doi:10.1038/ncb3024

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S., Göttgens, B., Blanpain, C., 2018. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science 359, 1177–1181. doi:10.1126/science.aao4174

Liu, F., Li, D., Yu, Y.Y.L., Kang, I., Cha, M.-J., Kim, J.Y., Park, C., Watson, D.K., Wang, T., Choi, K., 2015. Induction of hematopoietic and endothelial cell program orchestrated by ETStranscription factor ER71/ ETV2. EMBO Rep 16, 654–669. doi:10.15252/embr.201439939

Lun, A.T.L., Bach, K., Marioni, J.C., 2016. Pooling across cells to normalize single-cell RNA

sequencing data with many zero counts. 17, 75–14. doi:10.1186/s13059-016-0947-7

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 17, 10. doi:10.14806/ej.17.1.200

Masumoto, H., Nakane, T., Tinney, J.P., Yuan, F., Ye, F., Kowalski, W.J., Minakata, K., Sakata, R., Yamashita, J.K., Keller, B.B., 2016. The myocardial regenerative potential of three-dimensional engineered cardiac tissues composed of multiple human iPS cell-derived cardiovascular cell lineages. Sci. Rep. 6, 29933–29910. doi:10.1038/srep29933

McCracken, I.R., Taylor, R.S., Kok, F.O., la Cuesta, de, F., Dobie, R., Henderson, B.E.P., Mountford, J.C., Caudrillier, A., Henderson, N.C., Ponting, C.P., Baker, A.H., 2019. Transcriptional dynamics of pluripotent stem cell-derived endothelial cell differentiation revealed by single-cell RNA sequencing. Eur. Heart J. 385, 9963117. doi:10.1093/eurheartj/ehz351

Meilhac, S.M., Lescroart, F., Blanpain, C., Buckingham, M.E., 2014. Cardiac cell lineages that form the heart. Cold Spring Harb Perspect Med 4, a013888–a013888. doi:10.1101/cshperspect.a013888

Mills, R., Titmarsh, D., Koenig, X., Parker, B., Ryall, J., Quaife-Ryan, G., Voges, H., Hodson, M., Ferguson, C., Drowley, L., Plowright, A., Needham, E., Wang, Q., Gregorevic, P., Xin, M., Thomas, W., Parton, R., Nielsen, L., Launikonis, B., James, D., Elliott, D., Porrello, E., Hudson, J., 2017. Functional screening in human cardiac organoids reveals a metabolic mechanism for cardiomyocyte cell cycle arrest. Proc. Natl. Acad. Sci. U.S.A. 114(40), E8372-E8381. doi:10.1073/pnas.1707316114

Moretti, A., Caron, L., Nakano, A., Lam, J.T., Bernshausen, A., Chen, Y., Qyang, Y., Bu, L., Sasaki, M., Martin-Puig, S., Sun, Y., Evans, S.M., Laugwitz, K.-L., Chien, K.R., 2006. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. Cell 127, 1151–1165. doi:10.1016/j.cell.2006.10.029

Morita, R., Suzuki, M., Kasahara, H., Shimizu, N., Shichita, T., Sekiya, T., Kimura, A., Sasaki, K.-I., Yasukawa, H., Yoshimura, A., 2015. ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. Proc. Natl. Acad. Sci. U.S.A. 112, 160–165. doi:10.1073/pnas.1413234112

Narmoneva, D.A., Vukmirovic, R., Davis, M.E., Kamm, R.D., Lee, R.T., 2004. Endothelial cells promote cardiac myocyte survival and spatial reorganization: implications for cardiac regeneration. Circulation 110, 962–968. doi:10.1161/01.CIR.0000140667.37070.07

Ng, E.S., Davis, R., Stanley, E.G., Elefanty, A.G., 2008. A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nat Protoc 3, 768–776. doi:10.1038/nprot.2008.42

Paffett-Lugassy, N., Singh, R., Nevis, K.R., Guner-Ataman, B., O'Loughlin, E., Jahangiri, L., Harvey, R.P., Burns, C.G., Burns, C.E., 2013. Heart field origin of great vessel precursors relies on nkx2.5-mediated vasculogenesis. Nat Cell Biol 15, 1362–1369. doi:10.1038/ncb2862

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., Wu, S.M., Red-Horse, K., Quertermous, T., Wu, J.C., 2018. Large-Scale Single-Cell RNA-Seq Reveals Molecular Signatures of Heterogeneous Populations of Human Induced Pluripotent Stem Cell-Derived Endothelial Cells. Circ. Res. 123, 443–450. doi:10.1161/CIRCRESAHA.118.312913

Palpant, N.J., Pabon, L., Rabinowitz, J.S., Hadland, B.K., Stoick-Cooper, C.L., Paige, S.L., Bernstein, I.D., Moon, R.T., Murry, C.E., 2013. Transmembrane protein 88: a Wnt regulatory protein that specifies cardiomyocyte development. Development 140, 3799–3808. doi:10.1242/dev.094789

Quaranta, R., Fell, J., Rühle, F., Rao, J., Piccini, I., Araúzo-Bravo, M.J., Verkerk, A.O., Stoll, M., Greber, B., 2018. Revised roles of ISL1 in a hES cell-based model of human heart chamber specification. Elife 7, 12209. doi:10.7554/eLife.31706

Ravenscroft, S.M., Pointon, A., Williams, A.W., Cross, M.J., Sidaway, J.E., 2016. Cardiac Non-myocyte Cells Show Enhanced Pharmacological Function Suggestive of Contractile Maturity in Stem Cell Derived Cardiomyocyte Microtissues. Toxicol. Sci. 152, 99–112. doi:10.1093/toxsci/kfw069

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. doi:10.1093/bioinformatics/btp616

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J.I., Inoue, T., 1999. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 126, 3437–3447.

Stevens, K.R., Kreutziger, K.L., Dupras, S.K., Korte, F.S., Regnier, M., Muskheli, V., Nourse, M.B., Bendixen, K., Reinecke, H., Murry, C.E., 2009. Physiological function and transplantation of scaffold-free and vascularized human cardiac muscle tissue. Proc Natl Acad Sci U S A 106, 16568–16573. doi:10.1073/pnas.0908381106

Tam, P.P., Parameswaran, M., Kinder, S.J., Weinberger, R.P., 1997. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. Development 124, 1631–1642.

Tirziu, D., Giordano, F.J., Simons, M., 2010. Cell Communications in the Heart. Circulation 122, 928–937. doi:10.1161/CIRCULATIONAHA.108.847731

Tulloch, N.L., Muskheli, V., Razumova, M.V., Korte, F.S., Regnier, M., Hauch, K.D., Pabon, L., Reinecke, H., Murry, C.E., 2011. Growth of engineered human myocardium with mechanical loading and vascular coculture. Circ. Res. 109, 47–59. doi:10.1161/CIRCRESAHA.110.237206

van den Brink, S.C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C., van Oudenaarden, A., 2017. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat. Methods 14, 935–936. doi:10.1038/nmeth.4437

Vincent, S.D., Buckingham, M.E., 2010. How to make a heart: the origin and regulation of cardiac progenitor cells. Curr. Top. Dev. Biol. 90, 1–41. doi:10.1016/S0070-2153(10)90001-X

Whitfield, M.L., George, L.K., Grant, G.D., Perou, C.M., 2006. Common markers of proliferation.

Nat. Rev. Cancer 6, 99–106. doi:10.1038/nrc1802

Wu, T.D., Nacu, S., 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26, 873–881. doi:10.1093/bioinformatics/btq057

Wu, T.D., Watanabe, C.K., 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859–1875. doi:10.1093/bioinformatics/bti310

Xu, T., Le, T.D., Liu, L., Su, N., Wang, R., Sun, B., Colaprico, A., Bontempi, G., Li, J., 2017. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics 33, 3131–3133. doi:10.1093/bioinformatics/btx378

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R., Flicek, P., 2016. Ensembl 2016. Nucleic Acids Res. 44, D710–D716. doi:10.1093/nar/gkv1157

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16, 284–287. doi:10.1089/omi.2011.0118

Zaffran, S., Kelly, R.G., Meilhac, S.M., Buckingham, M.E., Brown, N.A., 2004. Right ventricular myocardium derives from the anterior heart field. Circ. Res. 95, 261–268. doi:10.1161/01.RES.0000136815.73623.BE

Zhao, H., Choi, K., 2019. Single cell transcriptome dynamics from pluripotency to FLK1+ mesoderm. Development 146, dev.182097. doi:10.1242/dev.182097

Zhao, H., Choi, K., 2017. A CRISPR screen identifies genes controlling Etv2 threshold expression in murine hemangiogenic fate commitment. Nat Comms 8, 541–512. doi:10.1038/s41467-017-00667-5

SUPPLEMENTARY INFORMATION

Supplemental figures S1-S7
Supplemental tables S1-S6
Supplemental Videos S1, S2
(Table S1-S4, Videos S1, S2 can be found online:
https://www.dropbox.com/sh/z96w86rp8insusk/AACzcYkq4eMufhe2UmQkia8Ca?dl=0)

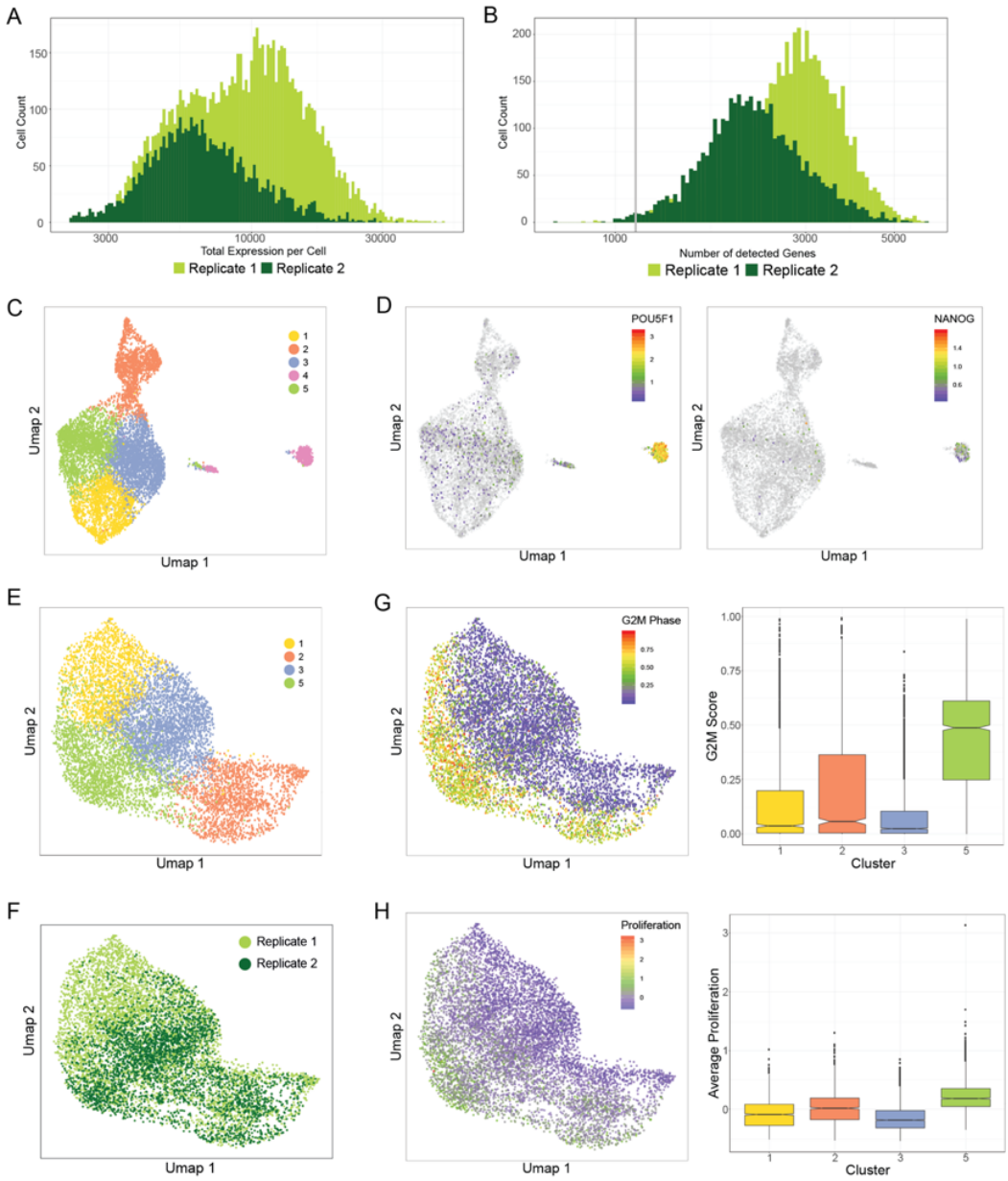

Figure S1. Single-cell RNA sequencing analysis for day 6 of CMEC differentiation from hiPSCs

(A-B) Distribution of total expression and number of detected genes in each cell of the scRNAseq dataset. Undetected genes (> 1 UMI count detected in less than two cells) and cells with low number of transcripts were removed from further analysis.

(C) scRNAseq data collected on day 6 is visualized using UMAP. Five clusters (1-5) of cells were identified.

(D) Expression of pluripotency markers *POU5F1* and *NANOG* in all cells shown in UMAP.

(E) UMAP after exclusion of pluripotent stem cell cluster (cluster 4).

(F) Overlay of two different batches of cells collected for scRNAseq on day 6 of CMEC

differentiation.

(G-H) G2M score (G) and standardized mean expression of proliferation-related genes (H) in all cells and in individual clusters.

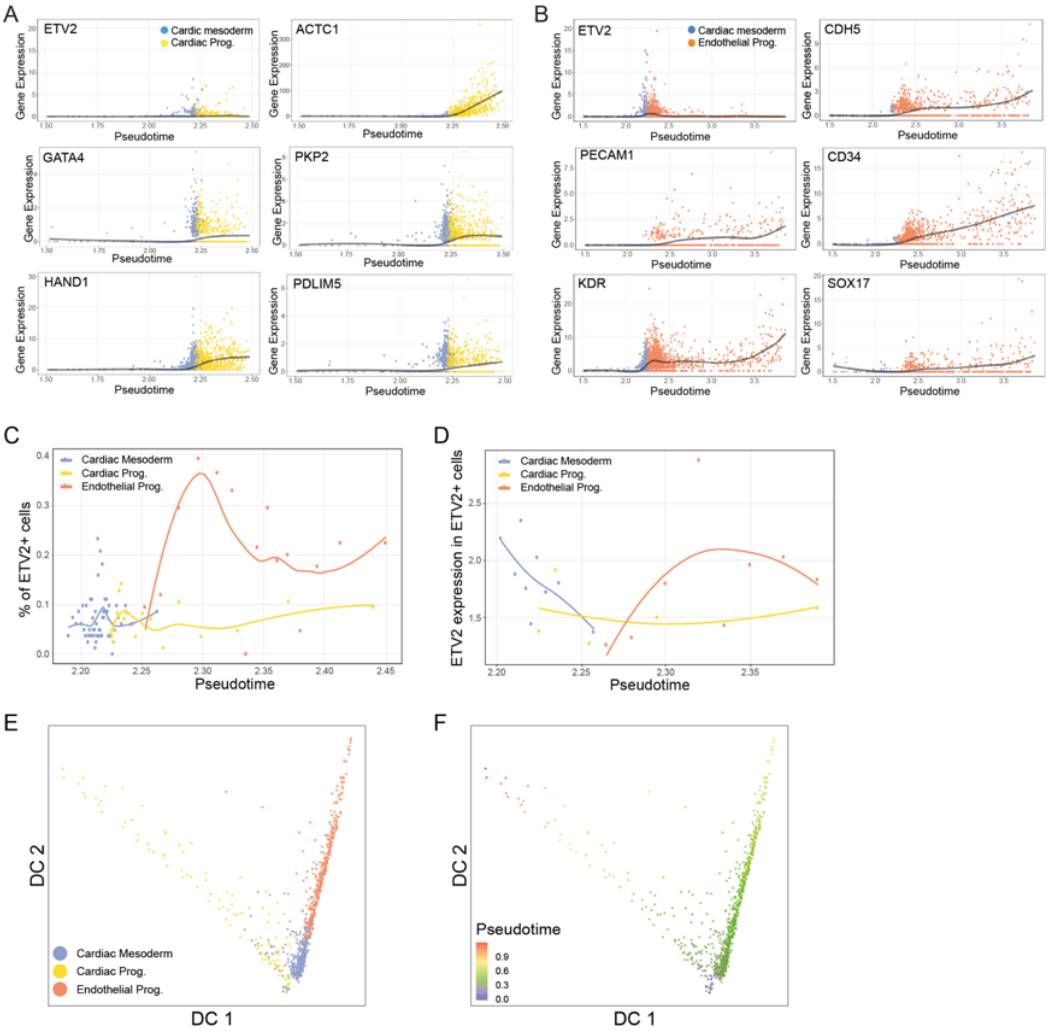

Figure S2. Pseudotime analysis of CMEC differentiation from hiPSCs

(A) Expression of *ETV2* and cardiac markers *ACTC1, GATA4, PKP2, HAND1* and *PDLIM5* following pseudotime in CM and CPs development route.

(B) Expression of *ETV2* and endothelial markers *CDH5, PECAM1, CD34, KDR* and *SOX17* following pseudotime in CM and EPs development route.

(C) Percentage of *ETV2* expressing cells within each branch between pseudotime 2.19 to 2.49. Bins of pseudo time with equal number of cells (50 cells) were constructed and the percentage of *ETV2* expressing cells was calculated among these cells.

(D) Average expression of *ETV2* in *ETV2* expressing cells within each branch between pseudotime 2.19 to 2.49. Bins of pseudo time with equal number of cells (50 cells) were constructed and the

average expression was calculated among these cells.

(E-F) Dimensionality reduction of scRNAseq data for all *ETV2*+ population using diffusion map. Cell identities are labeled with different colors in (E). Color scale indicates the pseudotime in (F).

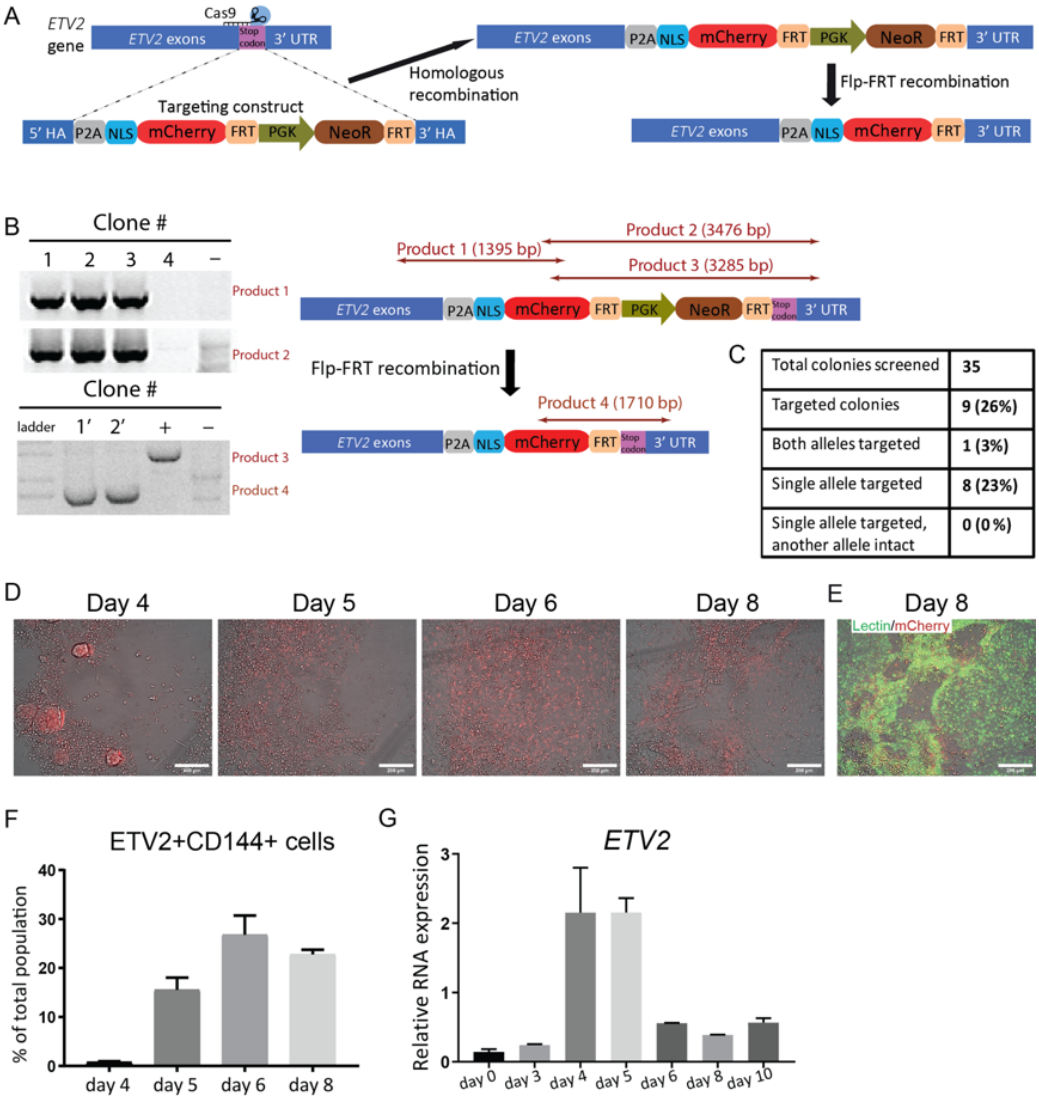

Figure S3. Generation and characterization of ETV2-mCherry reporter hiPSC line

(A) Schematic overview of CRISPR/Cas9-Mediated Knock-in of mCherry reporter into *ETV2* locus of hiPSCs. mCherry and Neomycin Resistance (NeoR) was inserted into *ETV2* gene locus through Homologous recombination. Then NeoR cassette was removed by flpo recombinase as shown on the right.

(B) PCR screening of targeted clones with correct insertion at *ETV2* locus. Two pairs of primers (for product 1 and product 2) were used to confirm the integration of construct into the genome.

Clone 1, 2, 3 were correctly targeted and clone 4 was not targeted. non-targeted hiPSCs (-) were included as negative control. Lower panel: Excision of neomycin-resistance cassette was confirmed using one pair of PCR primer (product 3 and 4 for before and after excision). Clone 1' ad 2' were successfully excised by FLPO recombinase. Genomic DNA before excision (+) and non-targeted hiPSCs (-) were included as positive and negative control.

(C) Summary of CRISPR targeting efficiency. Out of 35 colonies screened, 1 colony was targeted on both alleles. 8 colonies were targeted on only one allele, but the other allele showed undesired mutations.

(D) Overlay of bright-field and mCherry fluorescence images on differentiation day 4, 5, 6 and 8 of CMEC differentiation using ETV2-mCherry reporter line. Scale bar 200 µm.

(E) Lectin staining of cells at the same location as shown in (B) on day 8 of differentiation. mCherry is in red and Lectin is in green. Scale bar 200 µm.

(F) Quantification of percentages of ETV2+CD144+ cells by FACS on differentiation day 4, 5, 6 and 8.

(F) Quantification of *ETV2* expression by qPCR on different days of CMEC differentiation.

Error bars are ± SD of three independent experiments in (F-G).

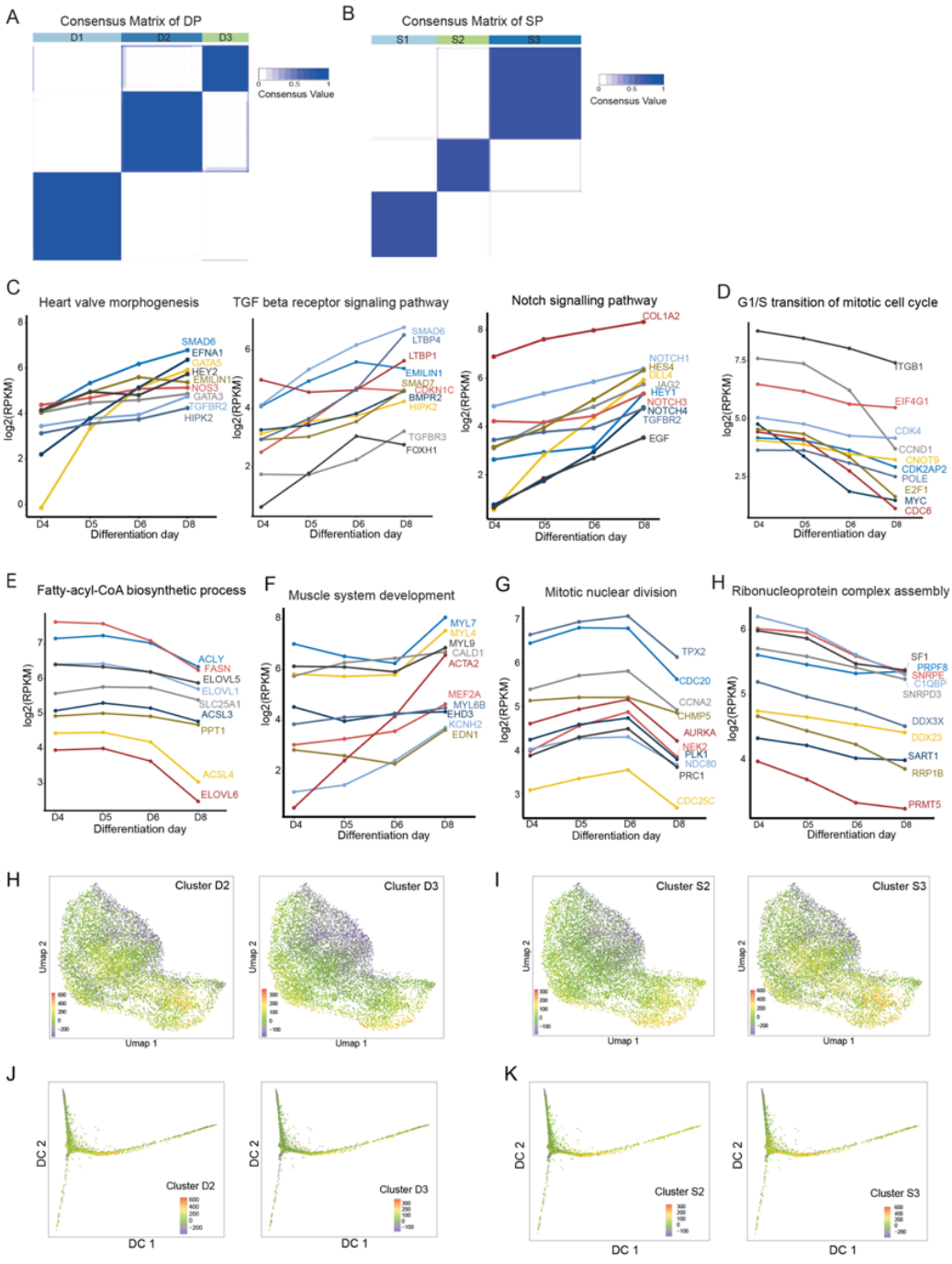

Figure S4. Time-course transcriptomic analysis of DP and SP cells during CMEC differentiation

(A) Consensus clustering for 3000 most variant genes across all DP samples (A) and all SP samples (B). All genes were divided into 3 clusters D1-D3 for DP (A) and S1-S3 for SP (B). Consensus value indicates similarity between two genes.

(C-H) Representative GOs enriched in cluster D1 (C), D2 (D), D3 (E), S1 (F), S2 (G) and S3 (H). Representative genes mapped to these GOs and their expression levels from day 4 to day 8 are

shown.

(H-I) Standardized sum of expression of genes in cluster D2, D3 (H) and S2, S3 (I) was calculated for each cell based on scRNAseq data and visualized in UMAPs. Color represents standardized sum of expression value.

(J-K) Standardized sum of expression of genes in cluster D2, D3 (J) and S2, S3 (K) was calculated for each cell based on scRNAseq data and displayed in diffusion maps. Color represents standardized sum of expression value.

Antibody	Application	Source	Dilution	Catalog #
CD144	FACS	eBioscience	1:50	53-1449-42
ICAM1	FACS	R&D	1:20	BBA20
E-Selectin	FACS	R&D	1:20	BBA21
α-ACTININ	IF	Sigma–Aldrich	1:800	A7811
cTnT	IF			
CD31	IF	R&D	1:200	AF806
ZO-1	IF	eBioscience	1:200	61-7300
CD144	IF	CellSignaling	1:200	2158
SM22	IF	Abcam	1:400	ab14106
AF488	IF	Invitrogen	1:200	A21206
AF594	IF	Termo Fisher	1:200	
AF647	IF	Termo Fisher	1:200	A21448

Table S4. List of antibodies

Gene	Forward sequence	Reverse sequence	Product size
hARP	CACCATTGAAATCCTGAGTGATGT	TGACCAGCCCAAAGGAGAAG	116
RPL37A	GTGGTTCCTGCATGAAGACAGTG	TTCTGATGGCGGACTTTACCG	84
ETV2	CAGCTCTCACCGTTTGCTC	AGGAACTGCCACAGCTGAAT	106
CDH5	GGCATCATCAAGCCCATGAA	TCATGTATCGGAGGTCGATGGT	100
CD31	GCATCGTGGTCAACATAACAGAA	GATGGAGCAGGACAGGTTCAG	101
PDGFRA	ATTGCGGAATAACATCGGAG	GCTCAGCCCTGTGAGAAGAC	95
NKX2-5	TTCCCGCCGCCCCCGCCTTCTAT	CGCTCCGCGTTGTCCGCCTCTGT	138
TBX5	ACATGGAGCTGCACAGAATG	TGCTGAAAGGACTGTGGTTG	104
GATA4	GACAATCTGGTTAGGGGAAGC	GAGAGATGCAGTGTGCTCGT	105
BMP10	CCTCTGCCAACATCATTAGGAG	TTTTCGGAGCCCATTAAAACTGA	77
HAND2	ACATCGCCTACCTCATGGAC	TGGTTTTCTTGTCGTTGCTG	162

Table S5. Sequence of primes used for qPCR