



Universiteit  
Leiden

The Netherlands

## Unravelling cell fate decisions through single cell methods and mathematical models

Mircea, M.

### Citation

Mircea, M. (2022, December 20). *Unravelling cell fate decisions through single cell methods and mathematical models*. *Casimir PhD Series*. Retrieved from <https://hdl.handle.net/1887/3505763>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3505763>

**Note:** To cite this publication please use the final published version (if applicable).

# 4

## TISSUE MICROENVIRONMENT PARTIALLY REMOVES SIGNATURES OF DEVELOPMENTAL ORIGIN IN A 3D IN VITRO MODEL OF CARDIAC ENDOTHELIAL CELL DIFFERENTIATION

4

*Endothelial cells (ECs) are ubiquitous across different organs of the body. They line the surface of different types of vessels, forming a tight barrier between a liquid and the surrounding tissue, and carry out various crucial functions. Dysfunction of ECs can therefore lead to a wide range of diseases. Importantly, the functions of ECs are highly organ-specific. For example, cardiac ECs line the blood vessels in the heart and have an essential role in nutrient transport. It is currently unknown, how ECs obtain their organ-specific characteristics. Developmental origin and tissue microenvironment are two possible determining factors, but their relative importance is unclear. This study uses human-induced pluripotent stem cells to derive ECs from two developmental origins: paraxial and cardiac mesoderm. We compare their characteristics and further integrate them into a cardiac microtissue, containing cardiomyocytes and fibroblasts, to investigate the influence of the surrounding cells on EC identity. Upon integration into the microtissue, the developmental origin is partially removed, and ECs from both developmental origins acquire an intramyocardial signature.*

---

This chapter is based on Xu Cao\*, **Maria Mircea\***, Francijna E. van den Hil, Hailiang Mei, Konstantinos Anastasiadis, Christine L. Mummery, Stefan Semrau and Valeria V. Orlova. Tissue microenvironment partially removes signatures of developmental origin in a 3D in vitro model of cardiac endothelial cell differentiation. Manuscript to be submitted. (\*contributed equally)

## 4.1 INTRODUCTION

All organs of the body rely on highly specialized cells to carry out their particular functions. Yet, there are cells with similar characteristics that appear across many different organs. A prime example are endothelial cells (ECs). ECs line the interior surfaces of vessels and form a tight barrier between a liquid, such as lymph or blood, and the surrounding tissue [1–6]. Due to their ubiquity, EC dysfunction is associated with a wide range of disease states, including atherosclerosis, diabetes, heart failure, hypertension and ischemia [7]. Despite the similarity in function, ECs from different vessels (artery, capillary, vein, lymphatic) show significant differences, reflecting the specific requirements of a particular environment. It is currently not well understood, how ECs (or other ubiquitous cell types) acquire their organ-specific characteristics. Developmental origin likely plays a role but cues from the microenvironment could be equally important. Using cardiac ECs as a model, this study explores the factors that lead to organ-specific molecular profiles.

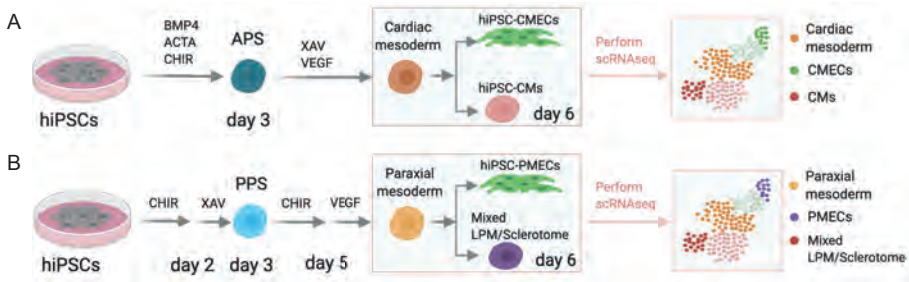
In the adult heart, ECs constitute > 60% of nonmyocytes [8] and can be divided into two subtypes: endocardial ECs (eEC) and intramyocardial ECs (iECs). eECs form the innermost layer and serve as a barrier between blood and myocardium. iECs constitute the coronary vessels, which transport oxygen and nutrients to the heart. In the mouse heart, eECs and iECs are distinguished by specific expressions of the eEC markers *Nfatc1*, *Npr3*, *Tmem100*, *Cdh11*, *Hapln1* [9–12], and the iEC markers *Apln*, *Fabp4*, *Cd36* [9, 13–15]. Equivalent EC subtypes have also been identified in the human heart, by single-cell RNA sequencing (scRNA-seq) [16–18]. Some of the eEC and iEC markers identified in the mouse are conserved in humans, such as *CDH11*, *NPR3* in eECs, and *CD36*, *FABP4* in iECs [17]. Unlike eECs, which have a sole origin, iECs have been shown to have multiple origins. The first established origin is the proepicardium on the venous pole of the heart, which later migrates onto the heart and gives rise to the epicardium and a small number of iECs [13, 19, 20]. More recently, sinus venosus [13, 16] and endocardium [10, 15] were revealed as two additional, major sources of iECs, by lineage tracing in mice. Postnatally, the endocardium continues to give rise to iECs that generate the majority of coronary vessels in the myocardium closest to the endocardium [21].

In addition to distinct developmental origins, the cardiac microenvironment could also endow eECs and iECs with tissue specific signatures. eECs are formed initially in the heart tube and exposed to blood flow, while iECs are developed first as a vascular plexus in the myocardium and are surrounded by other cell types like cardiomyocytes and fibroblasts. Cardiomyocytes are known to produce a vast amount of vascular endothelial growth factor-A (VEGF-A) [22] and other factors like Angiopoetin-1 [23], nitric oxide (NO), endothelin-1 (ET-1), fibroblast growth factor (FGF)-2, urocortin, haemoxygenase and adenosine [24], which are all key regulators of the EC phenotype. Other cardiac cell types, including fibroblast, vascular smooth muscle cells (VSMCs) and macrophages, also affect heart EC functions through either direct cell-cell contact or paracrine factors [25].

Most of what we know about EC development has been established in the mouse. Human *in vitro* systems provide a convenient, easily manipulated model to study human EC development. Human induced pluripotent stem cells (hiPSCs) provide an unlimited source of ECs [26, 27], which can be differentiated through different mesodermal origins including lateral plate mesoderm and paraxial mesoderm (PM) [28]. Several studies derived organ-

specific ECs from hiPSCs. Two groups claimed to have obtained brain microvascular ECs from hiPSCs through either co-differentiation or co-culture with neuronal cells [29, 30]. Our group previously developed a method to co-differentiate cardiomyocytes and ECs from hiPSCs through cardiac mesoderm (CM). These hiPSC-derived ECs expressed a number of cardiac specific genes like MEOX2, GATA4, GATA6 and ISL136, while tissue specific identities (of eEC or iEC) were still absent.

*In vitro* systems also allow probing the influence of the microenvironment, because different cell types can be easily combined. Recently our group established heart microtissues (MTs), a 3D cell culture model composed of ECs, cardiomyocytes and fibroblasts, all derived from hiPSCs [31, 32]. This model provides an ideal tool to investigate the influence of both the developmental origin and the cardiac specific microenvironment on the acquisition of a tissue specific EC identity. In this study, we derived ECs from hiPSCs through two mesodermal origins (CM and PM), using our established protocol [31] and an adapted protocol from the literature [28]. MTs were then generated using these two sources of ECs respectively. Interestingly, although newly differentiated ECs from two origins showed distinct identities, they became more similar after extended culture in MTs. Furthermore, based on eEC- and iEC-specific signatures extracted from a published single-cell RNA sequencing (scRNA-seq) dataset of human fetal heart [18], we observed an iEC rather than an eEC identity for both developmental origins after MT culture. In summary, this study shows that, although certain characteristics are inherited from progenitors, ECs efficiently adapt to the microenvironment and acquire new tissue-specific signatures. Our results provide new insights into the acquisition of organ/tissue-specific cell identities, which will inform the preparation of hiPSC-derived, organ specific ECs for disease modeling and drug development.



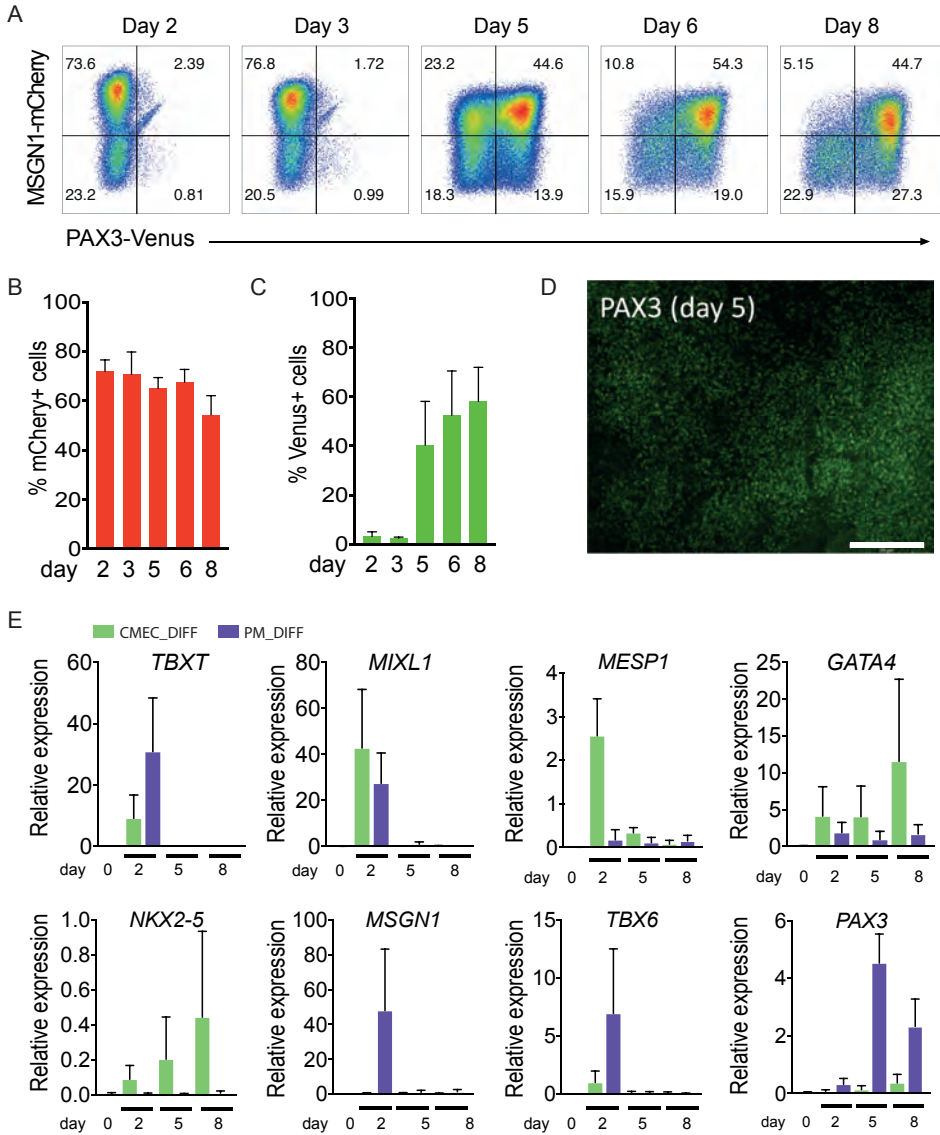
**Figure 4.1: Single-cell RNA sequencing analysis of endothelial cells differentiated from cardiac and paraxial mesoderm(A-B)** Schematic overview of CMEC (A) and PMEC (B) differentiation protocols until day 6. Cells were collected for scRNAseq on day 6. ACTA: activin-A. CHIR: CHIR99021. APS/PPS: anterior/posterior primitive streak. LPM: lateral plate mesoderm.

## 4.2 RESULTS

### 4.2.1 DERIVATION OF ENDOTHELIAL CELLS FROM DIFFERENT MESODERMAL ORIGINS

We set out to derive ECs from hiPSCs via two different mesodermal intermediates. To obtain cardiac mesoderm-derived ECs (CMECs), we used a protocol that was established previously in our group [31] (Figure 4.1A). Briefly, BMP4, Activin A (ACTA) and CHIR99021 (CHIR) were used to induce anterior primitive streak (APS) on day 3. Then, CMECs and early cardiomyocytes were induced on day 6 in the presence of XAV-939 (XAV) and VEGF. To obtain paraxial mesoderm-derived ECs (PMECs), we developed a new protocol (Figure 4.1B) based on a published approach [28]. Briefly, posterior primitive streak was derived through induction by CHIR for two days and subsequently XAV for one day. Next, PMECS and a mixed lateral plate mesoderm/sclerotome population were derived by exposure to CHIR for two days and VEGF for one day.

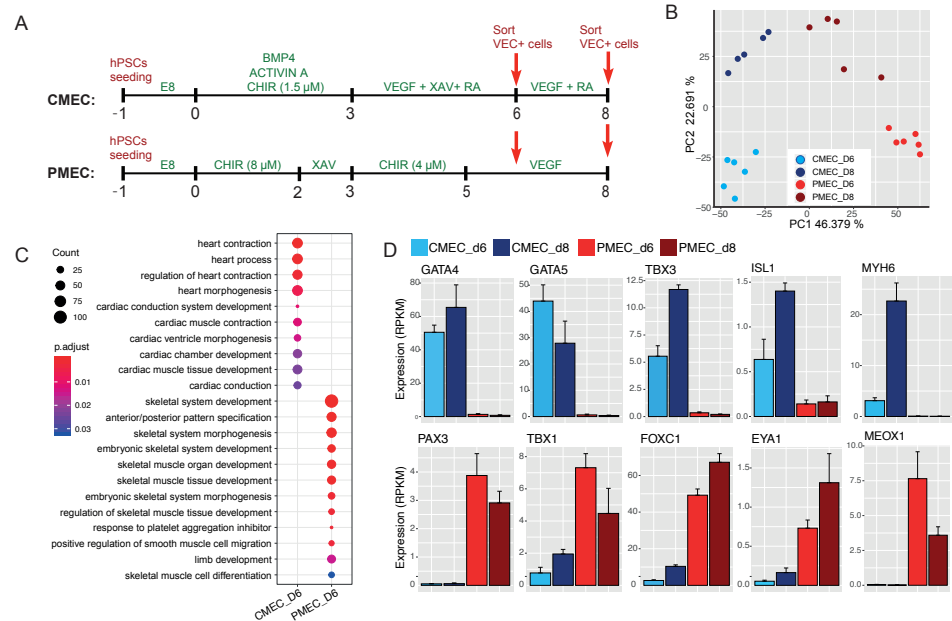
To facilitate the characterization of the newly developed PMEC protocol, we generated an hiPSC line harboring fluorescent reporters for PAX3 and *MSGN1* (*PAX3<sup>Venus</sup> MSGN1<sup>mCherry</sup>*) using CRISPR/Cas9 and the piggyBac transposon system. More than 70% of the cells expressed *MSGN1<sup>mCherry</sup>* on day 2 of PMEC differentiation (Figure 4.2 A, B). *MSGN1* expression persisted in 50% of the cells until day 8. However, *MSGN1* mRNA was only detectable on day 2 (Figure 4.2E). On day 5, around half of the cells started to express *PAX3<sup>Venus</sup>* (Figure 4.2A,C). Most of these cells also expressed *MSGN1<sup>mCherry</sup>*. Both *PAX3<sup>Venus</sup>* and *PAX3* mRNA were highly expressed from day 5 to day 8 (Figure 4.2A, C-E). On day 2, pan-mesoderm markers *TBXT* and *MIXL1* were expressed in both the CMEC and the PMEC differentiation protocol. While cardiac genes (*MESP1*, *GATA4* and *NKX2-5*) were exclusively expressed in CMEC differentiation, paraxial mesoderm related genes (*MSGN1*, *TBX6*, *PAX3*) were specifically expressed during PMEC differentiation (Figure 4.2E). Taken together, mRNA measurements and assessment of reporter fluorescence suggested that our new protocol produces ECs with paraxial mesoderm characteristics.



**Figure 4.2: Characterization of PMEC differentiation using *MSGN1<sup>mCherry</sup> PAX3<sup>Venus</sup>* dual reporter line**(A) FACS analysis of *PAX3<sup>Venus</sup>* and *MSGN1<sup>mCherry</sup>* expression at day 2, 3, 5, 6 and 8 of PMEC differentiation. (B-C) Quantification of percentages of mCherry+ (B) and Venus+ (C) cells in the total population by flow cytometry on day 2, 3, 5, 6 and 8. (D) Representative fluorescence image of *PAX3<sup>Venus</sup>* expression on day 5 of PMEC differentiation. Scale bar represents 200  $\mu$ m. (E) Quantification of *TBXT*, *MIXL1*, *MESP1*, *GATA4*, *NKX2-5*, *MSGN1*, *TBX6* and *PAX3* expression by qPCR on day 0, 2, 5 and 8 of CMEC (green) and PMEC (purple) differentiation.

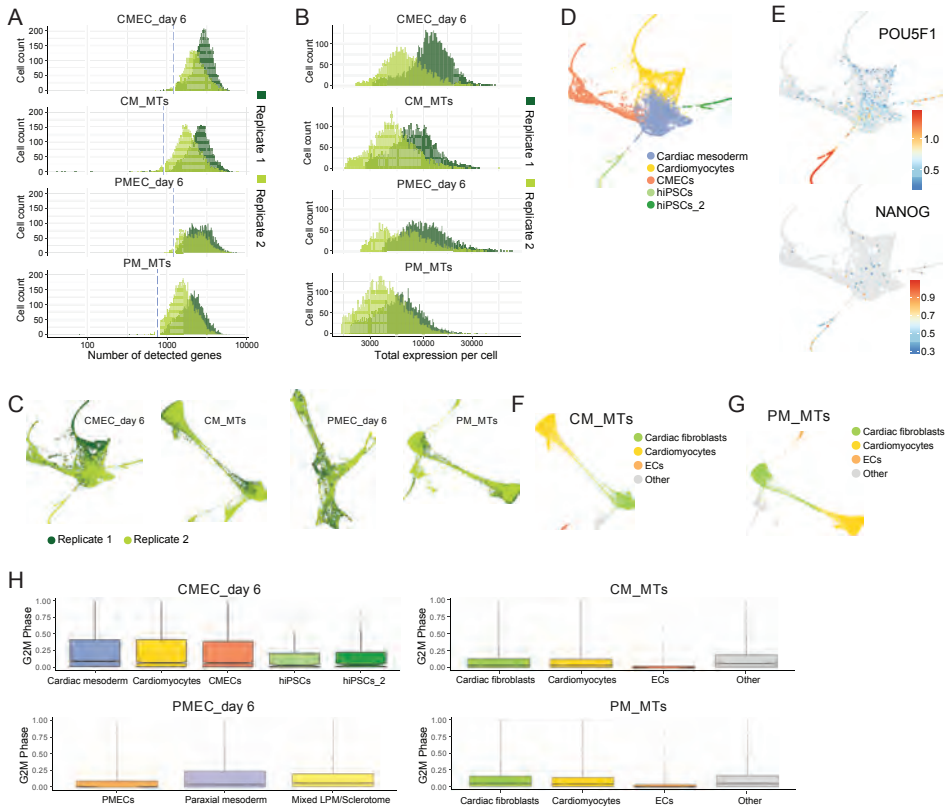
## 4.2.2 TRANSCRIPTOMIC PROFILING OF CMECS AND PMECS

To compare CMECs and PMECs more broadly, cells expressing the EC marker VEC were sorted on day 6 and day 8 of both protocols and characterized by bulk RNA sequencing (RNA-seq) (Figure 4.3A). Principle component analysis (PCA) showed that CMECs and PMECs clustered separately along PC1, and day 6 and day 8 were separated along PC2 (Figure 4.2B). On day 6, 3307 and 2592 genes were significantly differentially upregulated ( $p_{adjusted} \leq 0.05$ , fold-change  $\geq 2$ ) in CMECs and PMECs respectively (Table S1). Gene ontology (GO) analysis showed that cardiac related genes were specifically upregulated in day 6 CMECs (CMEC\_D6), while genes related to skeletal system development and function were specifically upregulated in day 6 PMECs (PMEC\_D6) (Figure 4.2C, Table S2). Genes involved in heart development, like GATA4, GATA5, TBX3, ISL1 and MYH6, were highly expressed in day 6 and day 8 CMECs. TBX3, ISL1 and MYH6 were upregulated from day 6 to day 8 in CMECs. Essential genes for skeletal muscle development like PAX3, TBX1, FOXC1, EYA1 and MEOX1 were majorly expressed in day 6 and day 8 PMECs. FOXC1 and EYA1 were upregulated from day 6 to day 8, while TBX1 and MEOX1 were downregulated (Figure 4.2D). In summary, unbiased expression analysis by bulk RNA-seq confirmed the respective mesodermal origins of the two derived EC cultures.



**Figure 4.3: Characterization of hiPSC-ECs differentiated using CMEC and PMEC protocols**(A) Schematic overview of CMEC and PMEC differentiation protocols from day -1 to day 8. VEC+ cells were sorted on day 6 and 8 from both protocols for bulk RNAseq. (B) PCA analysis of hiPSC-ECs sorted on day 6 and 8 of CMEC and PMEC protocol. (C) GO enrichment analysis for differentially expressed genes between CMECs and PMECs on day 6 of differentiation. The Complete list of GO terms can be found in Table S2. Color represents the enrichment p-value adjusted for multiple hypothesis testing and dot size represents the number of genes mapped to the GO term. (D) Normalized gene expression levels (RPKM) of cardiac and skeletal-related genes in CMECs and PMECs on day 6 and 8. Error bars indicate standard deviation.

### 4.2.3 CHARACTERIZATION OF CMEC AND PMEC DIFFERENTIATION BY SINGLE-CELL RNA-SEQ



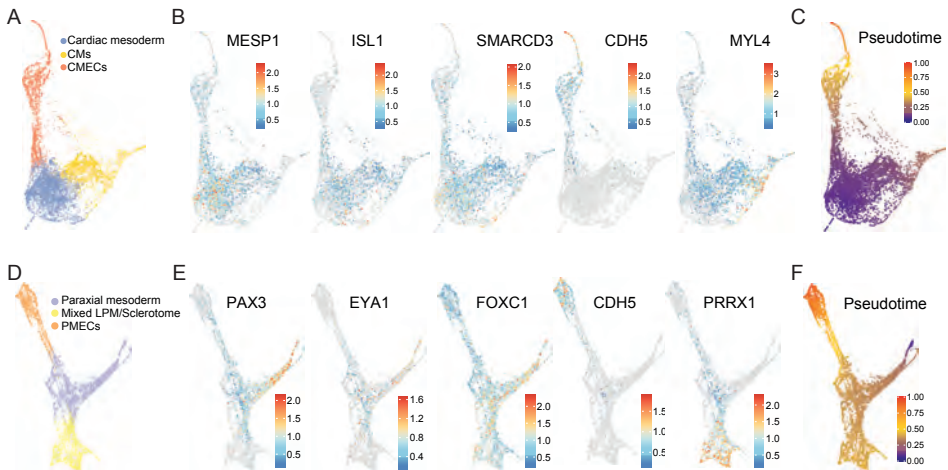
**Figure 4.4: Quality control of scRNAseq datasets (A-B)** Distribution of number of detected genes (A) and total expression (B) in each cell of the scRNAseq datasets. Dotted blue line indicates quality control threshold for datasets. Two different batches are labelled with different colors. (C) Two different batches of cells collected for each scRNAseq dataset are visualised using force-directed graph layout. (D) scRNAseq data of CMECs on day 6 is visualized using PAGA. Five cell clusters were identified and labelled with different colors. (E) Expression of pluripotency genes POU5F1 and NANOG in the CMEC dataset on day 6 is shown in PAGA plot. Color represents log transformed expression value. (F-G) scRNAseq data of CM\_MTs (F) and PM\_MTs (G) are visualized using PAGA. Four cell clusters were identified. Cluster labelled with “Other” was excluded in the downstream analysis in both datasets. (H) Boxplot of G2M phase-score in individual clusters of each dataset.

To reconstruct the differentiation trajectories of ECs, single-cell RNAseq (scRNA-seq) was performed on day 6 of CMEC and PMEC differentiation for two independent biological replicates (Figure 4.1, A-B, Figure 4.4). The replicates appeared highly similar in a low-dimensional representation (Figure 4.4C) and were therefore combined for further analysis. Undifferentiated hiPSCs remaining in the culture were excluded from further analysis (Figure 4.4D-E). In the CMEC differentiation data set, cells were grouped in 3 clusters, which we identified as cardiac mesoderm, cardiomyocytes and CMECs by marker gene analysis (Figure 4.5A, Table S3). The cardiac mesoderm cluster was characterized by mesoderm and early cardiac genes, such as MESP1, SMARCD3, ABLIM1, TMEM88, ISL1, MYL5, as



well as the cell cycle-related genes *CDK6* and *NEK2*. The CMEC cluster was characterized by EC markers (*CDH5*, *CD34*, *KDR*, *HEY2*, *TEK*, *TIE1*, *ACVRL1*, *SOX17*, *ENG*, *ICAM2*, *PECAM1*). Cardiomyocytes were identified by expression of cardiomyocyte-specific genes, including *MYL4*, *TNNI1*, *MYL7*, *ACTA2*, *TNNT2*, *HAND2* and *NKX2-5* (Figure 4.5B, 4.6A-B). Pseudo-time analysis showed that both CMECs and cardiomyocytes differentiated from cardiac mesoderm, and CMECs progressed further compared to cardiomyocytes (Figure 4.5C).

In the PMEC differentiation data set, all cells were divided into 3 clusters, which were interpreted as paraxial mesoderm, PMECs and mixed lateral plate mesoderm (LPM)/sclerotome (Figure 4.5D, Table S2) using marker gene analysis (Table S3). The paraxial mesoderm cluster was characterized by posterior primitive streak and dermomyotome genes, such as *MEOX1*, *PDGFRB*, *SIX1*, *CRABP2*, *NR2F1*, *EYA1*, *FOXC1* and *PAX3*. PMECs were characterized by EC markers, like *ETV2*, *CDH5*, *CD34*, *KDR*, *ENG*, *SOX17*, *PLVAP*, *APLN*, *NRP1*. The mixed LPM/sclerotome cluster was characterized by LPM and sclerotome specific genes, such as *TMEM88*, *HAND1*, *TNNI1*, *PRRX1*, *ACTA2*, *DES*, *FOXH1*, *LEF1* and *JAG1* (Figure 4.5E, 4.6C-D). Pseudo-time analysis showed that both PMECs and LPM/Sclerotome developed from paraxial mesoderm (Figure 4.5F). All in all, clustering and marker gene analysis of the scRNA-seq data confirmed the bulk RNA-seq results and revealed high similarity in developmental dynamics between the two differentiation protocols.



**Figure 4.5: Single-cell RNA sequencing analysis of endothelial cells differentiated from cardiac and paraxial mesoderm** (A) scRNAseq data of CMECs on day 6 is visualized using a force-directed graph layout (FGL). Three clusters of cells were identified. (B) FGL plots show expression (log transformed) patterns of *MESP1*, *ISL1*, *SMARCD3*, *CDH5*, *MYL4* in CMEC population. Color represents log transformed expression value. (C) Pseudotime analysis of CMEC dataset on day 6 of differentiation. (D) scRNAseq data of PMECs on day 6 is visualized using FGL. Three clusters of cells were identified. (E) FGL plots show expression (log transformed) patterns of *PAX3*, *EYA1*, *FOXC1*, *CDH5*, *PRRX1* in PMEC population. Color represents log transformed expression value. (F) Pseudotime analysis of PMEC dataset on day 6 of differentiation.

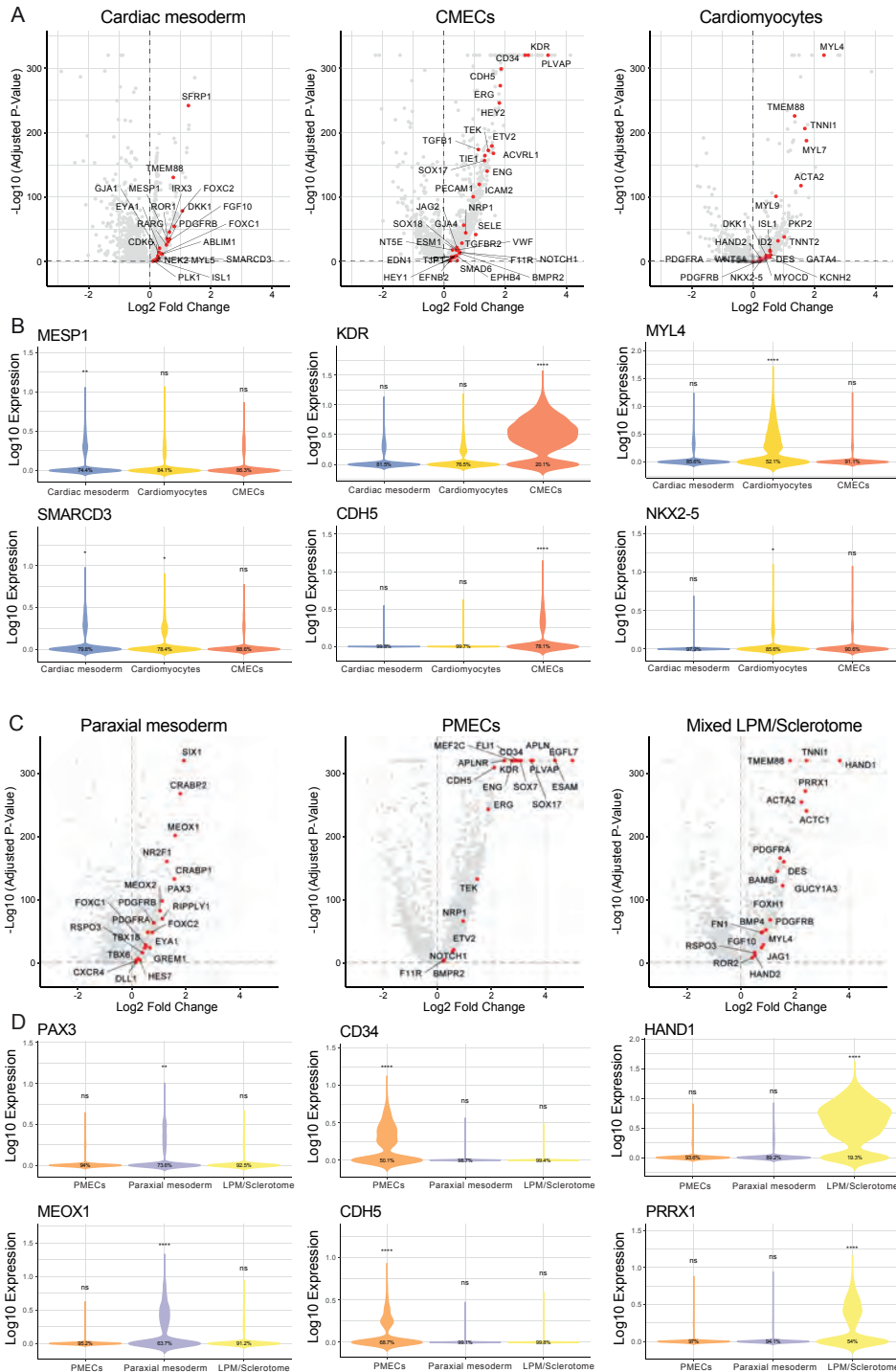


Figure 4.6: Single-cell RNA-sequencing analysis of CMEC and PMEC datasets on day 6 (Caption on the next page.)

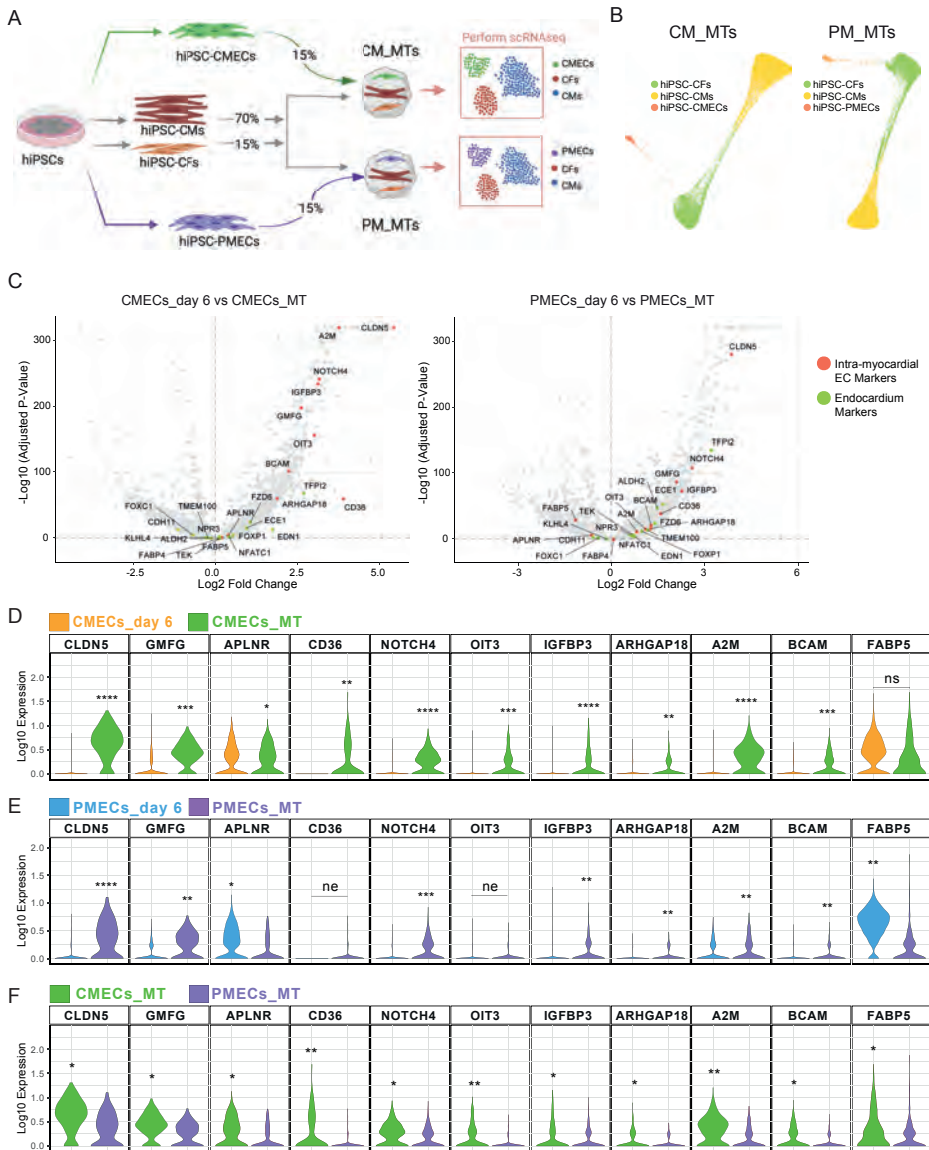
#### 4.2.4 hiPSC-ECs ACQUIRED ORGAN-SPECIFIC SIGNATURES IN A CARDIAC MICROENVIRONMENT

Being able to produce ECs with properties corresponding to their mesodermal origins finally enabled us to test in how far the cellular microenvironment can either reinforce or reverse this specification. Specifically, we set out to mimic the cardiac microenvironment *in vitro* using a protocol for creating cardiac microtissues (MTs), which was published previously by our group [32]. Briefly, CD34+ CMECs or PMECs were sorted on day 6 and combined with hiPSC-derived cardiomyocytes (hiPSC-CMs) and hiPSC-derived fibroblasts (hiPSC-CFs) at a ratio of 3:14:3 to form MTs. MTs made from CMECs (CM\_MTs) and PMECs (PM\_MTs) were collected after 21 days for scRNA-seq (Figure 4.7A). Two independent biological replicates were deemed highly similar (Figure 4.4C) and were therefore combined for further analysis. Both CM\_MTs and PM\_MTs datasets were divided into three clusters that correspond to hiPSC-CFs, hiPSC-CMs and hiPSC-ECs (Figure 4.7B). Marker genes identified for each cluster, confirmed the cluster identities (Table S6).

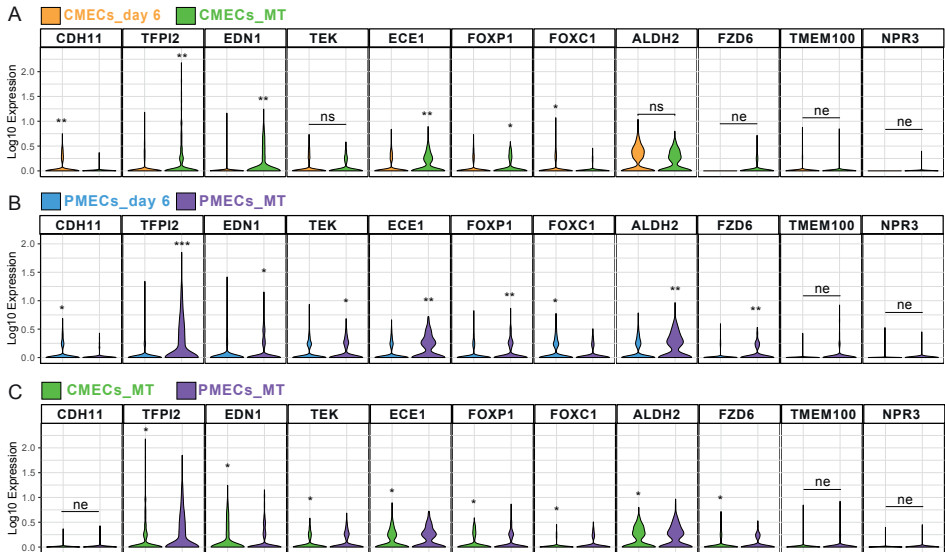
In both cases, a fourth cluster of cells with an uninterpretable signature was ignored (Figure 4.4F-G). Next, we compared the CMECs from monoculture differentiation on day 6 (CMEC\_day 6) with the CMECs in MTs (CMECs\_MT). Most intramyocardial makers including CLDN5, GMFG, APLNR, CD36, NOTCH4, OIT3, IGFBP3, ARHGAP18, A2M and BCAM and several endocardial markers (TFPI2, EDN1, ECE1, FOXP1, FOXC1) were upregulated in CMEC\_MT (Figure 4.7C-D, 4.8A, Table S7). However, the differences in endocardial marker expression were smaller compared to intramyocardial markers (Figure 4.7C-D, 4.8A, Table S7). Then, we compared PMECs from monoculture differentiation on day 6 (PMECs\_day 6) to PMECs in MTs (PMECs\_MT). Most intra-myocardial markers, including CLDN5, GMFG, NOTCH4, IGFBP3, ARHGAP18, A2M and BCAM, and some endocardium markers (TFPI2, EDN1, TEK, ECE1, FOXP1, ALDH2, FZD6) were upregulated in PMEC\_MT (Figure 4.7C, 4.8B, Table S7).

Compared to PMECs\_MT, CMECs\_MT expressed higher levels of intra-myocardial markers, especially APLNR, CD36, OIT3, ARHGAP18, A2M, BCAM which were barely expressed in the majority of PMECs\_MT cells (Figure 4.7F, Table S8). Although most endocardium markers were also higher in CMECs\_MT than in PMECs\_MT, their average expression levels were lower in general compared to intra-myocardial markers (Figure 4.8C, Table S8). Notably, most endocardial makers (including CDH11, FOXC1, FZD6, TMEM100 and NPR3) were barely expressed in both CMECs\_MT and PMECs\_MT, (Figure 4.8C). Overall, heart tissue-specific genes, especially intramyocardial markers, were upregulated in hiPSC-ECs upon extended culture in the cardiac microenvironment of the MT.

**Figure 4.6: scRNAseq analysis of CMEC and PMEC datasets on day 6 (Figure on previous page.)** (A) Volcano plots showing fold changes and adjusted p-values for differential gene expression between a specific cluster in the CMEC dataset and all other cells in that dataset. Representative, significantly up-regulated genes ( $P_{adjusted} \leq 0.05$  & fold change  $\geq 1.2$ ) are labelled in red. (B) MESP1, SMARCD3, KDR, CDH5, MYL4 and NKX2-5 expression (log transformed) in three clusters of CMEC dataset on day 6. (C) Volcano plots showing fold changes and p-values for differential gene expression between a specific cluster in the CMEC dataset and all other cells in that dataset. Representative significantly up-regulated genes ( $P_{adjusted} \leq 0.05$  & fold change  $\geq 1.2$ ) are labelled in red. (D) PAX3, MEOX1, CD34, CDH5, LEF1 and PRRX1 expression (log transformed) in three clusters of the PMEC dataset on day 6.



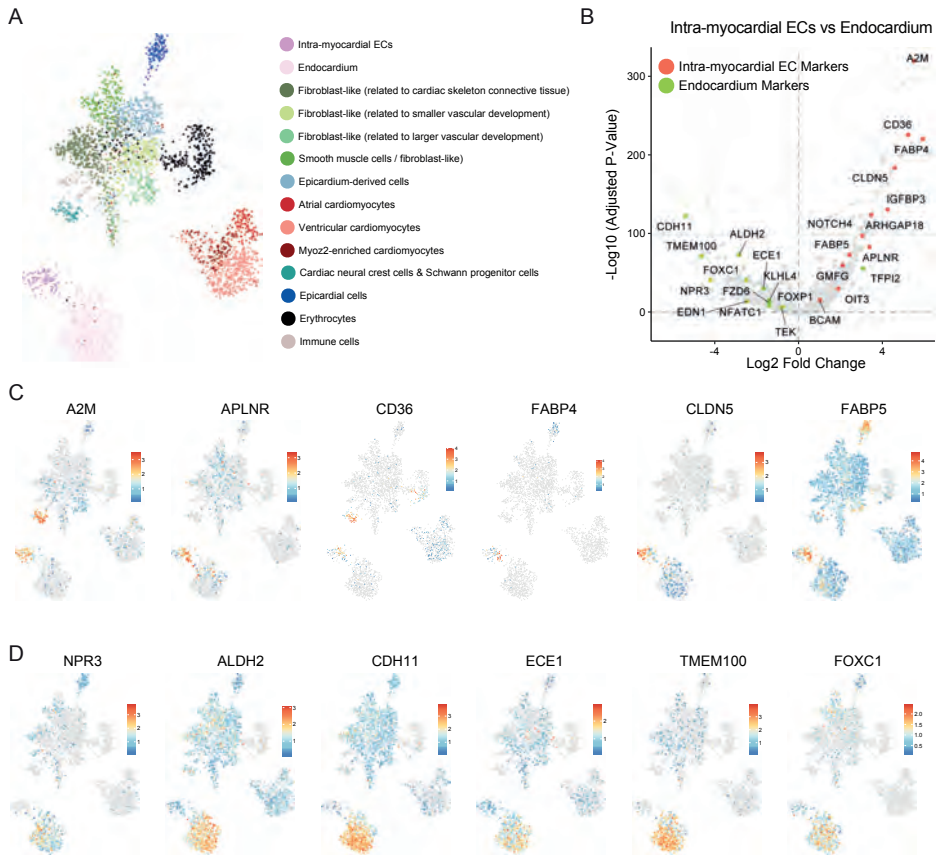
**Figure 4.7: hiPSC-ECs acquired organ-specific signatures in a cardiac microenvironment** (A) Schematic overview; generation of cardiac microtissues (MTs) from hiPSC-CMs, hiPSC-CFs and hiPSC-ECs. CMECs and PMECs were used for CM\_MTs and PM\_MTs, respectively. MTs were collected after 21 days for scRNAseq. (B) scRNAseq data of CM\_MTs (left) and PM\_MTs (right) were visualized using force-directed graph layout. Three clusters of cells were identified in both datasets. (C) Volcano plot shows fold changes and p values of all genes tested between two selected clusters: CMECs\_day 6, CMECs\_MTs (left), and PMECs\_day 6, PMECs\_MTs (right). Representative intra-myocardial and endocardial markers that are differentially expressed ( $P_{adjusted} \leq 0.05$ ) are labelled in red and green respectively. (D-F) Differential expression tests between CMECs\_day 6 and CMECs\_MT (D), PMECs\_day 6 and PMECs\_MT (E), CMECs\_MT and PMECs\_MT (F) for representative intra-myocardial EC markers. ns:  $p \geq 0.05$ ; \*  $p \leq 0.05$ ; \*\*  $p \leq 1e-10$ ; \*\*\*  $p \leq 1e-100$ ; \*\*\*\*  $p \leq 1e-200$ . Clusters with higher expression value were indicated with stars. ne: not expressed (0 counts) in  $\geq 85\%$  of cells in both groups.



**Figure 4.8: Characterization of endocardial signatures of hiPSC-ECs on day 6 and in MTs** (A-C) Differential expression tests between cluster CMECs\_day 6 and CMECs\_MT (C), PMECs\_day 6 and PMECs\_MT (D), CMECs\_MT and PMECs\_MT (E) for representative endocardial EC markers. ns:  $p \geq 0.05$ ; \*  $p \leq 0.05$ ; \*\*  $p \leq 1e-10$ ; \*\*\*  $p \leq 1e-100$ ; \*\*\*\*  $p \leq 1e-200$ . Clusters with higher expression value were indicated with stars. ne: not expressed (0 counts) in  $\geq 85\%$  of cells in both groups.

#### 4.2.5 EXTRACTION OF ORGAN-SPECIFIC SIGNATURES OF HUMAN FETAL HEART ECs FROM A PUBLISHED scRNA-SEQ DATASET

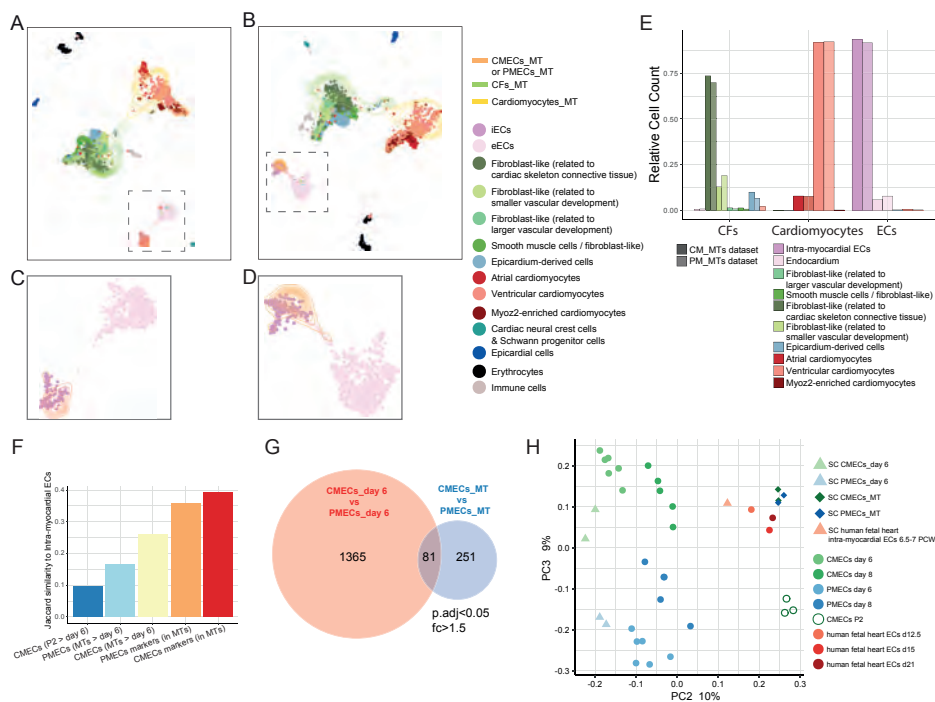
To assess how closely ECs in MTs assume an organ-specific identity, we sought to compare them to primary heart ECs. To that end, we re-analyzed a published scRNA-seq dataset of the fetal human heart (EGAS0000100399) [18] and identified the expression signature of heart ECs. The whole cell population was divided into 14 clusters. Our interpretation of these clusters deviated from the published annotation in two cases (Figure 4.9A): 1. The original endothelium/pericytes/adventitia cluster (cluster 10) was reannotated as intramyocardial ECs, based on differentially expressed markers such as A2M, CD36, APLNR, ARHGAP18, IGFBP3, CLDN5, FABP4 and FABP5. 2. The cluster annotated as capillary endothelium in the original publication (cluster 0) was reannotated as endocardium, due to the presence of differentially expressed markers like NPR3, ALDH2, CDH11, ECE1, TMEM100, FOXC1 and EDN1 (Figure 4.9B, Table S5). Supporting the differential expression test, UMAP visualization of representative intra-myocardial and endocardial markers showed specific expression in the respective clusters (Figure 4.9C-D). In conclusion, our clustering and differential expression analysis revealed distinct endothelial tissues in the fetal human heart.



**Figure 4.9: Re-analysis of a published scRNAseq dataset to identify organ specific signatures of human fetal heart ECs** (A) Dimensionality reduction of scRNAseq data of human fetal heart using Uniform MAnifold Approximation and Projection (UMAP). 14 cell clusters were identified and named based on their identities according to the original publication except for two EC clusters: intra-myocardial ECs and endocardium. (B) Volcano plot showing fold changes and adjusted p-values for differential gene expression between intra-myocardial ECs and endocardium. Representative differentially expressed genes ( $P_{adjusted} \leq 0.05$ ) that are known to be intra-myocardial or endocardial markers are labelled in red and green, respectively. (C-D) Expression (log transformed) of representative intra-myocardial EC markers (C) and endocardium markers (D) in individual cells (UMAPs).

#### 4.2.6 BOTH CMECS AND PMECS ACQUIRED INTRA-MYOCARDIAL IDENTITY IN MT CULTURE

To obtain a clear view of the similarities between ECs in MTs and primary fetal heart ECs, we combined the CM\_MT or PM\_MT dataset with the published *in vivo* data (Figure 4.10A-B). In both cases, hiPSC-CFs in MTs (CF\_MT) cluster together with fetal heart fibroblast-like cells; hiPSC-CMs in MTs (CM\_MT) cluster together with fetal heart ventricular cardiomyocytes. Notably, both CMECs\_MT and PMECS\_MT cluster together with fetal heart intra-myocardial ECs rather than endocardium (Figure 4.10A-D).



**Figure 4.10: Both CMECs and PMECs acquired intra-myocardial identity in a cardiac microenvironment** (A-B) The published human fetal heart scRNAseq dataset (EGAS0000100399) was combined with the CM\_MTs (A) or PM\_MTs (B) dataset and visualized using Uniform Manifold Approximation and Projection (UMAP). EC clusters are marked with dashed boxes. (C-D) Zoom-ins of EC clusters marked in (A-B), showing overlap of CMECs\_MT (C) and PMECs\_MT (D) with intramyocardial ECs of the fetal heart dataset. Cell clusters are labeled with different colors. Cells in the fetal heart dataset are represented with dots and cells in MT datasets are represented with contour lines. (E) The nearest clusters to the published dataset were identified for each cell of CM\_MTs and PM\_MTs. The result of knn-assignment is visualized in the bar plot. Cells from CM\_MTs and PM\_MTs are labeled in dark and light colors respectively. Identities of each cell clusters of the published dataset are shown with different colors. (F) Jaccard similarity to the marker genes of intra-myocardial ECs from the published fetal heart dataset was calculated for each group of genes. CMEC markers (in MTs): specific markers of CMECs within CM\_MT dataset; PMEC markers (in MTs): specific markers of PMECs within PM\_MT dataset; CMECs (MTs  $\geq$  day 6): differentially expressed genes (DEGs) that are higher in CMECs\_MT compared to CMECs\_day 6; PMECs (MTs  $\geq$  day 6): DEGs that are higher in PMECs\_MT compared to PMECs\_day 6; CMECs (P2  $\geq$  day 6): DEGs that are higher in passage two CMECs compared to CMECs\_day 6. (G) Venn diagram showing numbers and overlap of DEGs ( $P_{adjusted} \leq 0.05$  and  $foldchange \geq 1.5$ ) between CMECs and PMECs from day 6 (in red) and MTs (in blue). (H) PCA of different EC populations in scRNAseq (triangle and diamond) and bulk RNAseq (circle) datasets using the marker genes of the intra-myocardial ECs from the in vivo data set. Average expression values of all cells in the cluster were used for the scRNAseq data.

To quantify our observation, we calculated the distances (in expression space) between each cell in MTs and the fetal heart dataset. This calculation showed that CF\_MT cells are closest to fibroblast-like cells in vivo (related to cardiac skeleton connective tissue); CM\_MT cells are closest to ventricular cardiomyocytes; and CMECs\_MT as well as PMECs\_MT are closest to intra-myocardial ECs in vivo. Annotating the in vitro cells based on the closest in vivo neighbors revealed that cell type identities were very similar in CM\_MTs and PM\_MTs (Figure 4.10E). Correspondingly, the set of markers of the CMECs\_MT and

the PMECs\_MT cluster showed a high overlap (Jaccard similarity) with the markers of intra-myocardial ECs we extracted from the *in vivo* data set. The gene set upregulated in CMECs\_MT (compared to CMECs\_day 6) had a higher overlap with intra-myocardial EC markers than the set of genes upregulated in PMECs\_MTs (compared to PMECs\_day 6). Genes that were upregulated in passage two CMECs compared to CMECs\_day 6 overlapped the least with intra-myocardial EC markers (Figure 4.10F). CMECs\_MT and PMECs\_MT thus both resembled intra-myocardial ECs but a difference between the two differentiation systems remained. To quantify this difference directly, we used differential gene expression analysis. On day 6 of differentiation, 1446 genes were differentially expressed between CMECs and PMECs (Table S9), while only 332 genes were differentially expressed between CMECs\_MT and PMECs\_MT (Table S8). 81 genes were shared between the two sets (Figure 4.10G).

Next, all EC clusters from bulk and single cell RNA-seq datasets were combined and visualized using principal component analysis (PCA) (Figure 4.10H). CMECs\_day 6 and PMECs\_day 6 cluster far apart, while CMECs\_MT and PMECs\_MT clustered closely together. Bulk and single cell RNA-seq samples clustered together for both CMECs and PMECs. CMECs\_MT and PMECs\_MT were found close to fetal heart intra-myocardial ECs and fetal heart ECs sequenced in our previous study ([32])(Figure 4.10H). All in all, the integrated analysis of the *in vitro* and *in vivo* data sets revealed that the cardiac microenvironment, mimicked by cardiac MT, partially removed expression differences due to distinct mesodermal origins.

## 4.3 DISCUSSION

In this study we set out to delineate the possible factors that can confer organ-specific characteristics to cardiac ECs. In principle, those characteristics could be inherited from developmental precursors or induced by the tissue microenvironment. We investigated the contributions from both factors using hiPSC-derived ECs and cardiac MTs as tools.

To model different developmental origins, CMECs and PMECs were derived through CM and PM respectively. CMECs obtained a clear cardiac phenotype but no tissue-specific (eEC or iEC) identity. PMECs expressed a number of limb/skeletal muscle specific genes. Both protocols thus resulted in early or immature organ-specific identities. However, we cannot conclude that these identities were conferred by the respective mesoderm precursors in an entirely cell-autonomous manner, since other cell types were co-differentiated in both protocols and could have influenced the ECs. To exclude the contribution from other cells in the culture, early EC progenitors would have to be purified and further differentiated to establish whether organ-specific identities are present.

To model the influence of cell-extrinsic factors, we took advantage of our cardiac MT model which mimics the heart-specific microenvironment better than other *in vitro* models [3]. Both CMECs and PMECs acquired an iEC identity after incorporation into MTs and continued culture. This result supports that high plasticity [5, 6] allows ECs to adapt efficiently to signals from the microenvironment.

Notably, the iEC identity was more pronounced in CMECs\_MT than in PMECs\_MT, as it might take extra steps and therefore more time for PMECs to adopt a cardiac fate, compared to CMECs. Importantly, extended monoculture of CMECs did not result in an iEC identity,



suggesting that the heart-specific microenvironment, modeled by cardiac MTs, caused the specification. Some of the genes upregulated during MT culture are essential for EC function. For example, CLDN5 and NOTCH4 were among the most highly upregulated genes in both protocols. CLDN5 is critical for the tight junction and barrier functions of endothelium [8], while NOTCH signaling plays key roles in both vascular morphogenesis and adult endothelium homeostasis [9]. These results might indicate that ECs also mature functionally in MTs.

ECs have been incorporated into hPSC-derived organoids of liver [10, 11], intestine [12], brain [13–15], pancreas [16] and kidney [17, 18] by either codifferentiation or aggregation. In most studies, organ/tissue-specific identities were barely investigated. Camp et al. [33] used an approach similar to our MTs, combining hepatic, stromal, and endothelial cells to make liver organoids [11]. Interestingly, that study reported transcriptomic changes in ECs upon coculture in liver organoids, suggesting further maturation. Similarly, coculture with neuronal cells was found to endow hPSC-derived ECs with a brain microvascular identity [34]. Together with these previous studies, our results support the notion that the microenvironment plays an essential role in developing an organ/tissue-specific identity of ECs. More work is needed to investigate the molecular mechanisms underlying the influence of the microenvironment, which likely include direct cell-cell contact, secretion of paracrine factors and modulation of the extracellular matrix.

We hope that our findings will guide the derivation of organ-specific ECs in the future and lay the foundation for various biomedical applications.

## 4.4 MATERIALS AND METHODS

### 4.4.1 EXPERIMENTAL METHODS

#### hiPSC CULTURE

The NCRM1 hiPSC line (NIH) was used for all experiments in this study. The cells were 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, were grown on feeders in maintenance 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 *PAX3<sup>Venus</sup>MSGN1<sup>mCherry</sup>* hiPSC DUAL REPORTER LINE

The *PAX3<sup>Venus</sup>* reporter was introduced first by CRISPR/Cas9 as follows: NCRM1 hiPSCs were passaged using split ratios of 1:2 or 1:3. The cells were transfected in 60 mm dishes after reaching 60-70% confluency. For transfection, 20  $\mu$ l lipofectamine (Invitrogen), 8  $\mu$ g repair template and 8  $\mu$ g sgRNA/Cas9 plasmid were diluted in 600  $\mu$ l Opti-MEM and added to each 60 mm dish. After 18 h the medium was changed to maintenance medium. After another 6 h G-418 (50  $\mu$ g/ml) selection was started and kept for 1 week. Surviving cells were cultured in maintenance medium and grown in 6-well plates for the transfection with a Flp recombinase expression vector to remove the neomycin cassette. 300  $\mu$ l Opti-MEM containing 10  $\mu$ l lipofectamine and 4  $\mu$ g CAGGs-Flpo-IRES-puro plasmid was added per well for 18 h. Puromycin (0,5  $\mu$ g/ml) selection was started 24 h post transfection and lasted for 2 days. Once recovered, individual clones were expanded by limiting dilution in 96-well plates. Targeted clones were identified by PCR and Sanger sequencing (BaseClear). Next, the *MSGN1<sup>mCherry</sup>* reporter was integrated into the genome of the *PAX3<sup>Venus</sup>* reporter line using a piggyBac transposon system created by Katrin Neumann, Konstantinos Anastasiadis (Biotechnology Center TU Dresden).

#### ENDOTHELIAL CELL DIFFERENTIATION FROM hiPSCs

Endothelial cells with a cardiac mesoderm origin were induced from hiPSCs in a monolayer using the CMEC protocol as described previously [31]. Briefly hiPSCs were split at a 1:12 ratio and seeded in 6-well plates coated with 75  $\mu$ g/mL growth factor reduced Matrigel (Corning) on day -1. On day 0, cardiac mesoderm was induced by changing from TeSR-E8 medium to BPEL medium [35], supplemented with 20 ng/mL BMP4 (R&D Systems), 20 ng/mL ACTIVIN A (Miltenyi Biotec) and 1.5  $\mu$ M CHIR99021 (Axon Medchem). From day 3, the cells were grown in BPEL medium supplemented 5  $\mu$ M XAV939 (Tocris Bioscience) and 50 ng/ml VEGF (R&D Systems), with or without 100  $\mu$ M retinoic acid (MERCK) and the medium was refreshed every 3 days.

To derive endothelial cells with paraxial mesoderm origin, hiPSCs were split at a 1:12 ratio and seeded in 6-well plates coated with 75  $\mu$ g/mL growth factor reduced Matrigel on day -1. On day 0, paraxial mesoderm was induced by changing from TeSR-E8 to BPEL medium, supplemented with 8  $\mu$ M CHIR99021. On day 2, the medium was exchanged with BPEL medium supplemented with 5  $\mu$ M XAV939. On day 3, the medium was exchanged with BPEL medium supplemented with 4  $\mu$ M CHIR99021. From day 5 onwards, cells were grown in BPEL medium supplemented with 50 ng/ml VEGF and the medium was refreshed every 3 days.

### FLUORESCENCE-ACTIVATED CELL SORTING

Cells were dissociated with TrypLE on day 6 and 8 of the CMEC or PMEC protocol and stained with a VE-Cadherin (VEC) antibody conjugated with PE (R&D Systems). Then VEC-positive cells were purified using a FACSAria III cell sorter. Total RNA was extracted right after sorting using the NucleoSpin®RNA kit (Macherey-Nagel).

### GENERATION OF 3D CARDIAC MICROTISSUES (MTs)

Cardiac MTs were generated from hiPSC-derived ECs (hiPSC-ECs), hiPSC-derived cardiac fibroblasts (hiPSC-CFs) and hiPSC-derived cardiomyocytes (hiPSC-CMs) as previously described [32]. Briefly, on day 6 of CMEC or PMEC differentiation, CD34-positive hiPSC-ECs were isolated using a Human cord blood CD34 Positive selection kit II (StemCell Technologies) following the manufacturer's instructions. On the day of MT formation, freshly isolated hiPSC-ECs and cultured hiPSCs-CFs and hiPSC-CMs were combined together (70% hiPSC-CMs, 15% hiPSC-ECs and 15% hiPSCs-CFs) at a concentration of 5000 cells per 50 $\mu$ l in BPEL medium supplemented with VEGF (50 ng/ml) and FGF2 (5 ng/ml). Cell suspensions were seeded in V-bottom 96-well microplates (Greiner bio-one) and centrifuged for 10 min at 1100 rpm. MTs were incubated at 37 °C, 5% CO<sub>2</sub> for 21 days. The medium was refreshed every 3 to 4 days. single-cell RNA-sequencing analysis of MTs was performed after 21 days.

### FLOW CYTOMETRY ANALYSIS

Cells were dissociated with TrypLE, washed once with flow cytometry buffer (PBS containing 0.5% BSA and 2 mM EDTA) and analyzed on a 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).

### QUANTITATIVE REAL-TIME POLYMERASE CHAIN REACTION (qPCR)

Total RNA was extracted using the NucleoSpin®RNA kit 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  $\Delta$ CT calculation and normalized to housekeeping gene RPL37A.

### 4.4.2 BULK RNA SEQUENCING (RNA-SEQ) AND ANALYSIS

Bulk RNAseq of passage two CMECs (CMECs P2) and human fetal heart ECs from week 12.5 (W12.5), W15 and W21 of gestation were performed in our previous study [32] and obtained from GEO (accession number GSE116464).

Bulk RNAseq of day 6 and 8 of CMEC and PMEC differentiation were performed by BGI (Shenzhen, China) using the Illumina HiSeq4000 sequencer (100bp paired end reads). Raw data was processed using the LUMC BIOPET Gentrapp pipeline (<https://github.com/biopet/biopet>), which comprises FASTQ preprocessing, alignment and read quantification. Sickel (v1.2) was used to trim low-quality read ends (<https://github.com/najoshi/sickel>). Cutadapt (v1.1) was used for adapter clipping [36]. Reads were aligned to the human reference genome GRCh38 using GSNAP (gmap-2014-12-23) [37, 38] and htseq-count (v0.6.1p1) was for quantification using the Ensembl v87 annotation [39]. Biases related to gene length and GC content were corrected by conditional quantile normalization using the R package

cqn (v1.28.1) [40]. Genes were excluded if read count was below 5 in  $\geq 90\%$  of the samples. Differentially expressed genes were identified using a generalized linear model as implemented in edgeR (3.24.3) [41]. P-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure and q-values of  $\leq 0.05$  were considered significant. Analyses were performed using R (version 3.5.2). The Principal Component Analysis (PCA) plot was generated with the built-in R functions 'prcomp'. Spearman correlation between samples was calculated using the 'cor' function and the correlation heatmap was generated with a heatmap function from the NMF package. Gene ontology term enrichment was performed using the compareCluster function of the clusterProfiler package (v3.10.1) [42] and q-values  $\leq 0.05$  were considered significant.

### 4.4.3 SINGLE-CELL RNA SEQUENCING AND ANALYSIS

#### LIBRARY PREPARATION AND SEQUENCING

Cells were dissociated into single cells differentiation and loaded into the 10X Chromium Controller for library construction using the Single-Cell 3' Library Kit, Version 2 Chemistry (10x Genomics) according to the manufacturer's protocol. Next, indexed cDNA libraries were sequenced on the HiSeq4000 platform. Single-cell expression was quantified using unique molecular identifiers (UMIs) by 10x Genomics' "Cell Ranger" software.

Mean reads per cell for all eight data sets:

CMEC (R1): 28,499; CMEC (R2): 29,388; PMEC (R1): 31,860; PMEC (R2): 38,415; CM\_MT (R1): 39,319; CM\_MT (R2): 29,741; PM\_MT (R1): 36,726; PM\_MT (R2): 26,421.

#### SINGLE-CELL RNA-SEQUENCING DATA PRUNING AND NORMALIZATION

For data pruning and normalization, the two replicates of each of the 4 conditions (CMEC, PMEC, CM\_MT and PM\_MT) were combined without batch correction. Cells with a number of genes per cell below a certain threshold (1200 (CMEC), 1200 (PMEC), 900 (CM\_MT), 750 (PM\_MT), see Figure 4.4) were removed. Genes expressed in less than 2 of the remaining cells with a count of at most 1 were excluded from further analysis. Each combined data set was normalized using the R package scran (V 1.14.6) [43]. Highly variable genes (HVGs) were calculated (using 'improvedCV2' from the scran package) for each replicate of the combined data sets after excluding ribosomal genes, stress markers [44] and mitochondrial genes. For downstream analysis the top 5% HVGs were used after excluding proliferation [45] and cell cycle [46] related genes.

#### CELL CYCLE ANALYSIS AND BATCH CORRECTION

For each combined data set, cell cycle analysis was performed with the scran package using the 'cyclone' function [47] on normalized counts (Figure 4.4H). Cells with a G2/M score higher than 0.2 were considered to be in G2/M phase. Otherwise, they were classified as G1/S. Using this binary classifier as predictor, we regressed out cell cycle effects with the R package limma (V 3.42.2) [48] applied to log-transformed normalized counts. Then, for each combined data set, the two replicates were batch corrected with the fast mutual nearest neighbors correction method (MNN) [49] on the cell cycle corrected counts, using the 30 first principal components and 20 nearest-neighbors (Figure 4.4C).

## CLUSTERING

For each combined data set, batch-corrected counts were standardized per gene and then used to create a shared nearest neighbour (SNN) graph with the *scran* R package ( $d = 30$ ,  $k = 2$ ). Louvain clustering was applied to the SNN graph using the *igraph* python package (V 0.7.1) with these resolution parameters: 0.4 (CMEC), 0.4 (CM\_MT), 0.3 (PMEC), 0.1 (PM\_MT). For the CMEC data set, this resulted in 5 clusters (Figure 4.4D). Two of these 5 clusters were excluded from further analysis based on the expression of pluripotency markers (Figure 4.4E). For CM\_MT and PM\_MT, clustering resulted in 4 clusters (Figure 4.4F and 4.4G), where one cluster was excluded from further analysis, because it was mainly present in one of the two replicates. Additionally, the attempt to map this cluster to *in vivo* data resulted in mostly unassigned cell types (plot not shown). For PMEC, clustering resulted in 3 clusters.

## 4

### DIMENSIONALITY REDUCTION AND PSEUDOTIME

Dimensionality reduction was performed using the python *scanpy* pipeline (V 1.4.6). For both data sets (CMEC and PMEC) a 20 nearest-neighbors (knn,  $k=20$ ) graph was created from diffusion components of the batch corrected data sets. Diffusion components are the eigenvectors of the diffusion operator which is calculated from Euclidean distances and a Gaussian kernel. The aim is to find a lower dimensional embedding which considers the dynamics of differentiation. Both graphs were projected into two dimensions using a force-directed graph layout and starting positions obtained from the partition-based graph abstraction (PAGA) output [50]. PAGA estimates connectivities between partitions and performs an improved version of diffusion pseudotime. Diffusion pseudotime [50, 51] was calculated on these graphs with root cells selected from the “Cardiac Mesoderm” cluster in CMEC, and the “Paraxial Mesoderm” cluster in PMEC.

For CM\_MT and PM\_MT, the knn graphs ( $k=50$  for PM\_MT,  $k=100$  for CM\_MT) were created from the first 30 principal components of the batch-corrected data sets. These graphs were projected into two dimensions with a force-directed graph layout and starting positions from the PAGA output.

### IN VIVO DATA ANALYSIS AND MAPPING

The *in vivo* data set, downloaded from

<https://www.spatialresearch.org/resources-published-datasets/doi-10-1016-j-cell-2019-11-025/>, contains a 6.5 PCW (postcoitum weeks) human fetal cardiac tissue sample. The clusters and cluster annotations were obtained from the original publication [18]. The data set was normalized with the *scran* R package and HVGs were calculated as described above. Dimensionality reduction was performed with the R package *umap* (V 0.2.5.0) using 20 nearest-neighbors,  $\text{min\_dist} = 0.7$  and Euclidean distances.

### DIFFERENTIAL EXPRESSION ANALYSIS

All differential expression tests were performed with *edgeR* (V 3.28.1) [41] using a negative binomial regression and raw counts. The predictors in the regression were: cluster and replicate (both discrete variables), as well as the total number of counts per cell.

For marker gene analysis (Figures 4.6A and 4.6C), p-values were obtained for a contrast between the cluster of interest and all other clusters using regression coefficients averaged over the replicates. For tests between different data sets (Figure 4.7C), the corresponding

endothelial cell cluster was extracted from each data set. Then, a contrast between MT and day 6 was calculated by averaging over the predictors of both replicates. For the in vivo test (Figure 4.9B), intra-myocardial EC and endocardium clusters were extracted from the data set to calculate the contrast between them. P-values were adjusted for multiple hypothesis testing with the Benjamini-Hochberg procedure.

#### COMPARISON TO THE IN VIVO DATA SET

CM\_MT and PM\_MT data sets were mapped on the in vivo data set using the MNN method ( $d = 30$  principal components,  $k = 100$  nearest neighbors). First, in vitro replicates were mapped to each other, then the in vivo data was mapped on the combined in vitro data, using normalized, log-transformed counts and the 10% top HVGs of the in vivo data set. Dimensionality reduction was performed with the R package umap using 100 nearest-neighbors,  $\text{min\_dist} = 0.3$  and Euclidean distance.

K-nearest-neighbour (KNN) assignment was performed in the batch corrected, principal component space (30 PCs). The 100 nearest-neighbors in the in vivo data set based on Euclidean distances were calculated for each in vitro cell. The in vitro cell was ascribed the cell type most abundant among the 100 in vivo neighbors. Each such assignment received a confidence score, which is the number of in vivo neighbors with that cell type divided by the number of all nearest neighbors ( $=100$ ). A cell was not ascribed a cell type if either the average distance to its nearest neighbour exceeded a certain threshold (determined by the long tail of the histogram of average distances: 0.35), or the assignment had a confidence score smaller than 0.5. In addition, clusters containing less than 10 cells were not ascribed a cell type.

For the Jaccard similarity measure, marker genes of each differential expression test were selected with adjusted  $p\text{-value} \leq 0.05$ . The remaining genes were ranked by log<sub>2</sub> fold-change and the first 478 genes were selected for analysis. Then, the Jaccard distances were calculated between the marker genes of intra-myocardial endothelial cells and each of the other gene sets. For principal component analysis (Figure 4.10H), human fetal bulk samples [32] and in vitro bulk samples were combined with the single cell data sets. For each single-cell data set, the endothelial cells were extracted and the sum per gene over all cells was calculated. Then, bulk and single cell samples were log-transformed and combined into one data set. Principal component analysis was applied on the gene-wise standardized data set, using marker genes of the intra-myocardial endothelial cells from the in vivo data set.

#### 4.4.4 DATA AVAILABILITY

The accession numbers for the bulk and single-cell RNA-sequencing datasets reported in this paper are <https://www.ncbi.nlm.nih.gov/geo/>. Supplementary tables are available at <https://doi.org/10.5061/dryad.9p8cz8wkg>.

**Funding** This project received funding from the European Union's Horizon 2020 Framework Programme (668724); European Research Council (ERCAdG 323182 STEMCARDIO-VASC); 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). M. M. and S.S. were supported by the Netherlands Organisation for Scientific

Research (NWO/OCW, [www.nwo.nl](http://www.nwo.nl)), as part of the Frontiers of Nanoscience (NanoFront) program. The computational work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

**Disclosure of potential conflict of interest** The authors indicated no potential conflicts of interest.

**REFERENCES**

- [1] J. Kalucka et al. Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell*, 180(4):764–779.e20, feb 2020.
- [2] A. Jambusaria et al. Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation. *eLife*, 9, jan 2020.
- [3] R. Marcu et al. Human Organ-Specific Endothelial Cell Heterogeneity. *iScience*, 4:20, jun 2018.
- [4] D. J. Nolan et al. Molecular Signatures of Tissue-Specific Microvascular Endothelial Cell Heterogeneity in Organ Maintenance and Regeneration. *Developmental Cell*, 26(2):204–219, jul 2013.
- [5] M. Potente and T. Mäkinen. Vascular heterogeneity and specialization in development and disease. *Nature Reviews Molecular Cell Biology* 2017 18:8, 18(8):477–494, may 2017.
- [6] D. T. Paik et al. Single-Cell RNA Sequencing Unveils Unique Transcriptomic Signatures of Organ-Specific Endothelial Cells. *Circulation*, 142(19):1848–1862, nov 2020.
- [7] R. J. Esper et al. Endothelial dysfunction: a comprehensive appraisal. *Cardiovascular Diabetology*, 5:4, feb 2006.
- [8] A. R. Pinto et al. Revisiting cardiac cellular composition. *Circulation Research*, 118(3):400–409, 2016.
- [9] H. Zhang et al. Endocardium Minimally Contributes to Coronary Endothelium in the Embryonic Ventricular Free Walls. *Circulation Research*, 118(12):1880–1893, jun 2016.
- [10] B. Wu et al. Endocardial Cells Form the Coronary Arteries by Angiogenesis through Myocardial-Endocardial VEGF Signaling. *Cell*, 151(5):1083–1096, nov 2012.
- [11] S. Somekawa et al. Tmem100, an ALK1 receptor signaling-dependent gene essential for arterial endothelium differentiation and vascular morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(30):12064–12069, jul 2012.
- [12] M. C. Puri, J. Partanen, J. Rossant, and A. Bernstein. Interaction of the TEK and TIE receptor tyrosine kinases during cardiovascular development. *Development*, 126(20):4569–4580, oct 1999.
- [13] H. I. Chen et al. The sinus venosus contributes to coronary vasculature through VEGFC-stimulated angiogenesis. *Development (Cambridge)*, 141(23):4500–4512, dec 2014.
- [14] X. Tian et al. Subepicardial endothelial cells invade the embryonic ventricle wall to form coronary arteries. *Cell Research* 2013 23:9, 23(9):1075–1090, jun 2013.



- [15] K. Red-Horse, H. Ueno, I. L. Weissman, and M. A. Krasnow. Coronary arteries form by developmental reprogramming of venous cells. *Nature* 2010 464:7288, 464(7288):549–553, mar 2010.
- [16] H. Suryawanshi et al. Cell atlas of the foetal human heart and implications for autoimmune-mediated congenital heart block. *Cardiovascular Research*, 116(8):1446–1457, jul 2020.
- [17] Y. Cui et al. Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Reports*, 26(7):1934–1950.e5, feb 2019.
- [18] M. Asp et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell*, 179(7):1647–1660.e19, dec 2019.
- [19] R. E. Poelmann et al. Development of the cardiac coronary vascular endothelium, studied with antiendothelial antibodies, in chicken-quail chimeras. *Circulation Research*, 73(3):559–568, 1993.
- [20] T. C. Katz et al. Distinct Compartments of the Proepicardial Organ Give Rise to Coronary Vascular Endothelial Cells. *Developmental Cell*, 22(3):639–650, mar 2012.
- [21] X. Tian et al. Vessel formation. De novo formation of a distinct coronary vascular population in neonatal heart. *Science (New York, N.Y.)*, 345(6192):90–94, jul 2014.
- [22] F. J. Giordano et al. A cardiac myocyte vascular endothelial growth factor paracrine pathway is required to maintain cardiac function. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10):5780–5785, may 2001.
- [23] N. L. Ward et al. Angiopoietin 1 expression levels in the myocardium direct coronary vessel development. *Developmental Dynamics*, 229(3):500–509, mar 2004.
- [24] D. Tirziu, F. J. Giordano, and M. Simons. Cell Communications in the Heart. *Circulation*, 122(9):928–937, aug 2010.
- [25] F. Perbellini, S. A. Watson, I. Bardi, and C. M. Terracciano. Heterocellularity and Cellular Cross-Talk in the Cardiovascular System. *Frontiers in Cardiovascular Medicine*, 5:143, nov 2018.
- [26] V. V. Orlova et al. Functionality of endothelial cells and pericytes from human pluripotent stem cells demonstrated in cultured vascular plexus and zebrafish xenografts. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 34(1):177–186, jan 2014.
- [27] V. V. Orlova et al. Generation, expansion and functional analysis of endothelial cells and pericytes derived from human pluripotent stem cells. *Nature Protocols* 2014 9:6, 9(6):1514–1531, may 2014.
- [28] K. M. Loh et al. Mapping the pairwise choices leading from pluripotency to human bone, heart and other mesoderm cell-types. *Cell*, 166(2):451, jul 2016.

- [29] H. Minami et al. Generation of Brain Microvascular Endothelial-Like Cells from Human Induced Pluripotent Stem Cells by Co-Culture with C6 Glioma Cells. *PLOS ONE*, 10(6):e0128890, jun 2015.
- [30] E. S. Lippmann et al. Derivation of blood-brain barrier endothelial cells from human pluripotent stem cells. *Nature Biotechnology* 2012 30:8, 30(8):783–791, jun 2012.
- [31] E. Giacomelli et al. Three-dimensional cardiac microtissues composed of cardiomyocytes and endothelial cells co-differentiated from human pluripotent stem cells. *Development (Cambridge)*, 144(6):1008–1017, mar 2017.
- [32] E. Giacomelli et al. Human-iPSC-Derived Cardiac Stromal Cells Enhance Maturation in 3D Cardiac Microtissues and Reveal Non-cardiomyocyte Contributions to Heart Disease. *Cell Stem Cell*, 26(6), 2020.
- [33] J. G. Camp et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature* 2017 546:7659, 546(7659):533–538, jun 2017.
- [34] M. Buckingham, S. Meilhac, and S. Zaffran. Building the mammalian heart from two sources of myocardial cells. *Nature Reviews Genetics* 2005 6:11, 6(11):826–835, nov 2005.
- [35] E. S. Ng, R. Davis, E. G. Stanley, and A. G. Elefanty. 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. *Nature Protocols* 2008 3:5, 3(5):768–776, apr 2008.
- [36] M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12, may 2011.
- [37] T. D. Wu and C. K. Watanabe. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9):1859–1875, may 2005.
- [38] T. D. Wu and S. Nacu. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, 26(7):873–881, apr 2010.
- [39] A. Yates et al. Ensembl 2016. *Nucleic Acids Research*, 44(D1):D710–D716, jan 2016.
- [40] K. D. Hansen, R. A. Irizarry, and Z. Wu. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*, 13(2):204–216, apr 2012.
- [41] M. D. Robinson, D. McCarthy, Y. Chen, and G. K. Smyth. edgeR: differential expression analysis of digital gene expression data User’s Guide. *Bioinformatics*, 26(October 2018):1–75, 2013.
- [42] G. Yu, L. G. Wang, Y. Han, and Q. Y. He. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*, 16(5):284–287, may 2012.
- [43] A. T. Lun, K. Bach, and J. C. Marioni. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology*, 17(1):1–14, apr 2016.

- [44] S. C. van den Brink et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods*, 14(10):935–936, 2017.
- [45] M. L. Whitfield, L. K. George, G. D. Grant, and C. M. Perou. Common markers of proliferation. *Nature Reviews Cancer* 2006 6:2, 6(2):99–106, feb 2006.
- [46] B. Giotti, A. Joshi, and T. C. Freeman. Meta-analysis reveals conserved cell cycle transcriptional network across multiple human cell types. *BMC Genomics*, 18(1):1–12, jan 2017.
- [47] A. Scialdone et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods*, 85:54–61, sep 2015.
- [48] M. E. Ritchie et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, apr 2015.
- [49] L. Haghverdi, A. T. Lun, M. D. Morgan, and J. C. Marioni. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5):421–427, 2018.
- [50] F. A. Wolf et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20(1):1–9, mar 2019.
- [51] L. Haghverdi et al. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods* 2016 13:10, 13(10):845–848, aug 2016.