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

Mircea, M.

# 3

# ETV2 upregulation marks the specification of early cardiomyocytes and endothelial cells during co-differentiation



*The ability to differentiate human induced pluripotent stem cells (hiPSCs) efficiently into defined cardiac lineages, such as cardiomyocytes and cardiac endothelial cells, is crucial to study human heart development and model cardiovascular diseases in vitro. The mechanisms underlying the specification of these cell types during human development are not well-understood, which limits the fine-tuning and broader application of cardiac model systems. Here, we used the expression of ETV2, a master regulator of hematoendothelial specification in mice, to identify functionally distinct subpopulations during the co-differentiation of endothelial cells and cardiomyocytes from hiPSCs. Targeted analysis of single-cell RNA sequencing data revealed differential ETV2 dynamics in the two lineages. A newly created fluorescent reporter line allowed us to identify early lineage-restricted states and show that a transient ETV2-high state initiates the specification of endothelial cells. We further demonstrated that functional cardiomyocytes can originate, unexpectedly, from progenitors expressing ETV2 at a low level. Our study thus sheds light on the in vitro differentiation dynamics of two important cardiac lineages.*

This chapter is based on Xu Cao, **Maria Mircea**, Gopala Krishna Yakala, Francijna E. van den Hil, Hailiang Mei, Christine L. Mummery, Stefan Semrau, and Valeria V. Orlova. ETV2 upregulation marks the specification of early cardiomyocytes and endothelial cells during co-differentiation. Manuscript submitted.

## Significance statement

In vitro differentiation of cardiac cell types is of great importance for disease modeling and future regenerative medicine. Many of the relevant molecular mechanisms are currently not understood, which limits the efficiency and fine-tuning of existing differentiation protocols. Here, we focus on the master regulator ETV2 and show that its upregulation marks the specification of two cardiac cell types during co-differentiation. Using single-cell RNA-seq and a new fluorescent reporter line, we identify lineage-restricted subpopulations in the ETV2+ cells. Our study is the first to resolve ETV2 dynamics at the single-cell level in the context of in vitro human cardiac differentiation.



## 3.1 Introduction

In vivo, cardiomyocytes (CMs) and endothelial cells (ECs) originate from Mesp1+ progenitors specified during gastrulation. In mice, these cells appear in the primitive streak and subsequently migrate towards the lateral plate mesoderm around E6.5 [1–4]. It is still controversial when the segregation of CMs and ECs from their common progenitor occurs. Single-cell RNA-seq (scRNA-seq) of mouse Mesp1+ progenitors collected at E6.75 and E7.25 showed that these cells were already segregated into distinct cardiovascular lineages, including CMs and ECs [5]. However, other studies showed that multipotential progenitors were still present in Flk-1-expressing lateral plate mesoderm [6, 7]. These cells were the first to be recognized as multipotent cardiac progenitor cells (CPCs) [8]. Studies in mouse and chick showed that CPCs come from two different sources [9, 10]: the first and second heart field (FHF, SHF). The FHF in the cardiac crescent contributes to the primitive heart tube, which serves as a scaffold into which SHF cells can migrate before heart chamber morphogenesis. It has been shown that cells from the SHF are patterned before migration to give rise to different parts of the heart [2, 11]. CPCs from FHF and SHF can be distinguished by the expression of ISL1, which is specifically expressed in SHF [12]. NKX2-5 expressing CPCs in both FHF and SHF from E7.5 to E7.75 contribute to both CMs and ECs in the heart [13]. As a direct target of NKX2-5, ETV2 was found to be expressed in all ECs but not myocardium by E8.5 [14]. ETV2 was required for the development of endothelial and hematopoietic lineages and directly targets TAL1, GATA2, LMO2, TEK, NOTCH1, NOTCH4, and CDH5 [14–17]. In mouse embryonic stem cells (ESCs), VEGF-FLK1 signaling upregulates ETV2 expression to induce hemangiogenic specification via an ETV2 threshold-dependent mechanism [18]. ETV2 expression was also found to direct the segregation of hemangioblasts, and smooth muscle cells (SMCs) in mouse ESCs [19].
In human heart development, much less is known about the specification of endothelial and myocardial lineages from multipotent CPCs, both in terms of timing and gene regulatory mechanisms. More specifically, it is still unclear whether ETV2 also plays a role in the segregation of ECs and CMs from CPCs in humans. Overexpression of ETV2 converts human fibroblasts into endothelial-like cells [20] and ETV2 expression levels were manipulated in several studies to drive hiPSCs towards ECs in 2D, and 3D cultures [21–27]. Paik et al. performed scRNA-seq analysis of hiPSC-derived ECs (hiPSC-ECs), which made up less than 10% of the cells that expressed the cardiac maker TNNT2. The developmental dynamics of ECs and cardiac lineages as such were not further studied [28].

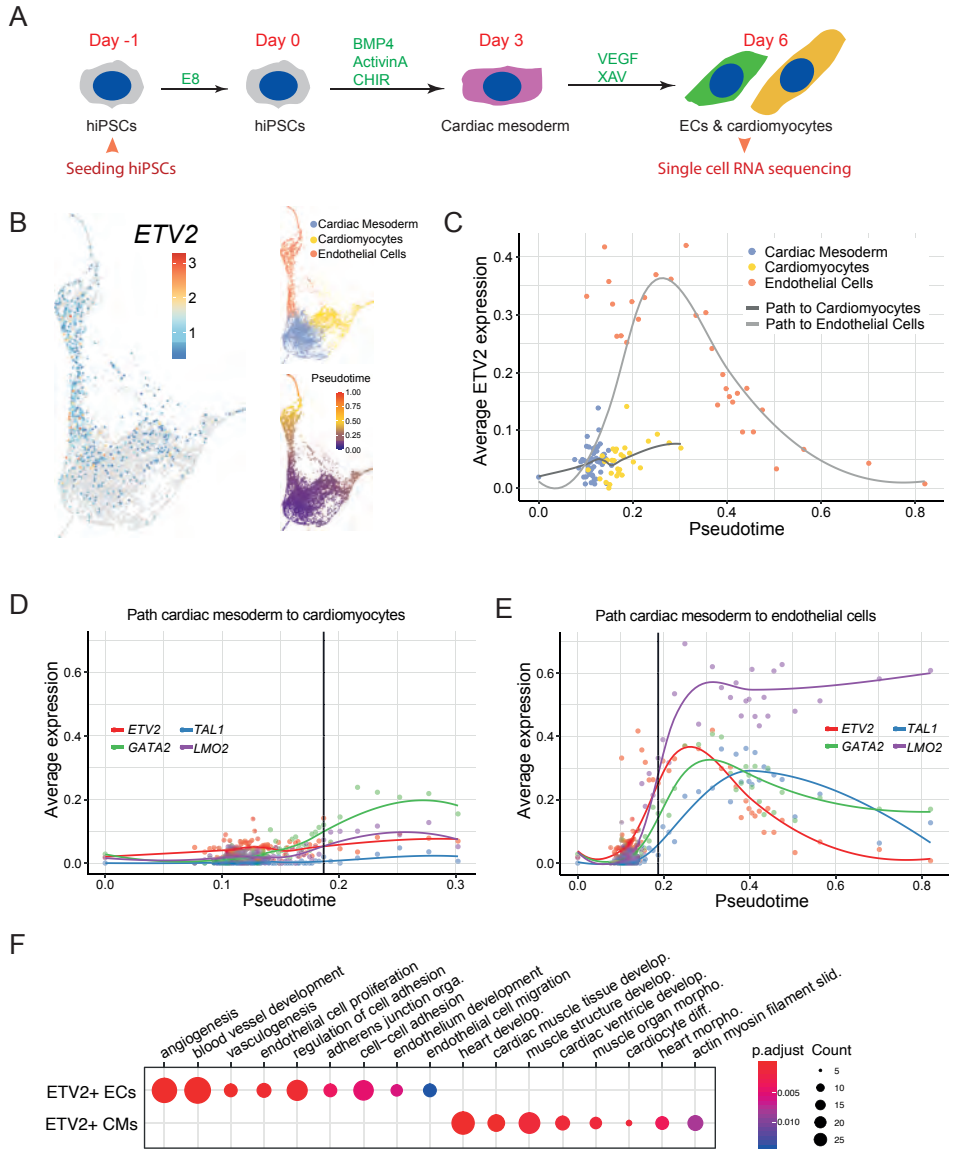


**Figure 3.1: scRNA-seq analysis of EC and CM co-differentiation reveals transient ETV2 upregulation after bifurcation.** (A) Schematic overview of the co-differentiation protocol from day -1 to day 6. Cells were collected for scRNA-seq on day 6. (B) Two-dimensional representation of the scRNA-seq data. Each data point is a single cell. Left: log2 transformed ETV2 expression is indicated by color. Top right: Cell identities are labeled with different colors. Bottom right: Pseudotime is indicated by color. (C) Average expression of ETV2 in bins of pseudotime for the developmental path of CMs or ECs . Cell identities are labeled with different colors. (D-E) Average expression of ETV2 and its direct target genes TAL1, GATA2, LMO across binned pseudotime along the developmental path of CMs (D) or ECs (E). Threshold (indicated in black) is set to the timepoint where the average ETV2 expression in EC reaches 0.25. (F) GO enrichment analysis of genes that were differentially expressed between ETV2+ ECs and ETV2+ CMs in the scRNA-seq dataset. 128 and 136 genes were upregulated in ETV2+ ECs and CMs respectively (Padjusted < 0.05). A complete list of GO terms can be found in Table S3. Color represents the Padjusted of the enrichment analysis and dot size represents the count of genes mapped to the GO term.

In a scRNA-seq study of hiPSC-ECs created with another differentiation protocol [29], ECs were collected at multiple time points. This study showed that endothelial and mesenchymal lineages had a common developmental origin in mesoderm cells, but the identity and differentiation potential of these cells was not characterized.
Previously, we found that MESP1+ progenitors derived from human ESCs could give rise to CMs, ECs and SMCs [30, 31]. We also developed a co-differentiation system for ECs and CMs from hiPSCs through a common cardiac mesoderm precursor [32]. Here we performed scRNA-seq analysis of this co-differentiation system to elucidate the relationship between ETV2 expression and specification of ECs and CMs from cardiac mesoderm. ETV2 expression was observed principally in the form of an initial pulse in the endothelial lineage but also in a subpopulation of the myocardial lineage. Using a newly generated $ETV2^{mCherry}$ hiPSC reporter line, we purified two subpopulations of ETV2+ cells and revealed their respective EC and CM expression characteristics by bulk RNA-seq. These sorted populations also showed distinct differentiation potentials towards CMs and ECs upon further differentiation with VEGF. In summary, this study depicted ETV2 dynamics during the segregation of human CMs and ECs differentiated from hiPSCs.

## 3.2 Results

### 3.2.1 ETV2 is upregulated after a bifurcation into CMs and ECs

To characterize the expression of ETV2 during co-differentiation of ECs and CMs [32] (Fig. 3.1A), we collected scRNA-seq data on day 6 of differentiation (Fig. 3.1B). We identified three distinct clusters: cardiac mesoderm, CMs, and ECs (Fig. 3.1B, top right and Table S3). Pseudotime analysis revealed cardiac mesoderm as the common developmental origin of CMs and ECs (Fig. 3.1B, bottom right). We found ETV2 to be highly expressed in the EC cluster, as well as a small fraction of cells in the cardiac mesoderm and CM clusters (Fig. 3.1B, left). We next focused on ETV2 expression dynamics along the developmental path from cardiac mesoderm to CMs and ECs. Notably, ECs extended to larger pseudotimes (0.15-0.8) compared to CMs (0.15-0.3), which might indicate faster differentiation dynamics in the EC lineage (Fig. 3.1C, 3.2A). After the bifurcation into ECs and CMs (around pseudotime 0.15), ETV2 increased only slightly in the CM lineage. In the EC lineage, however, it was initially strongly upregulated (until pseudotime 0.25), and subsequently declined to a similar level as in cardiac mesoderm (Fig. 3.1C). ETV2 downstream target genes, such as TAL1, GATA2, and LMO2 [17], were only slightly increased in the CM lineage (Fig. 3.1D), while in the EC lineage, they were highly induced and strongly correlated with ETV2 (Fig. 3.1E). Notably, TAL1, GATA2, and LMO2 only showed significant expression after ETV2 expression exceeded 0.25 in ECs, an expression level that was not reached in CMs (Fig. 3.2B-C). Endothelial specific genes KDR, CD34, SOX17, CDH5, and PECAM1 increased on the path from cardiac mesoderm to ECs (Fig. 3.2E). Most of these genes started to increase when ETV2 was already declining, as exemplified by CDH5 (Fig. 3.2D). Genes related to cardiac or muscle functionalities, like ACTC1, PDLIM5, HAND1, PKP2, and GATA4, most of which were already expressed in the cardiac mesoderm, were further increased in the CM lineage (Fig. 3.2F). Identification of genes that are differentially expressed between ETV2+ CMs and ECs showed enrichment in CM- and EC-specific genes, respectively (Fig. 3.1F, Table S4).

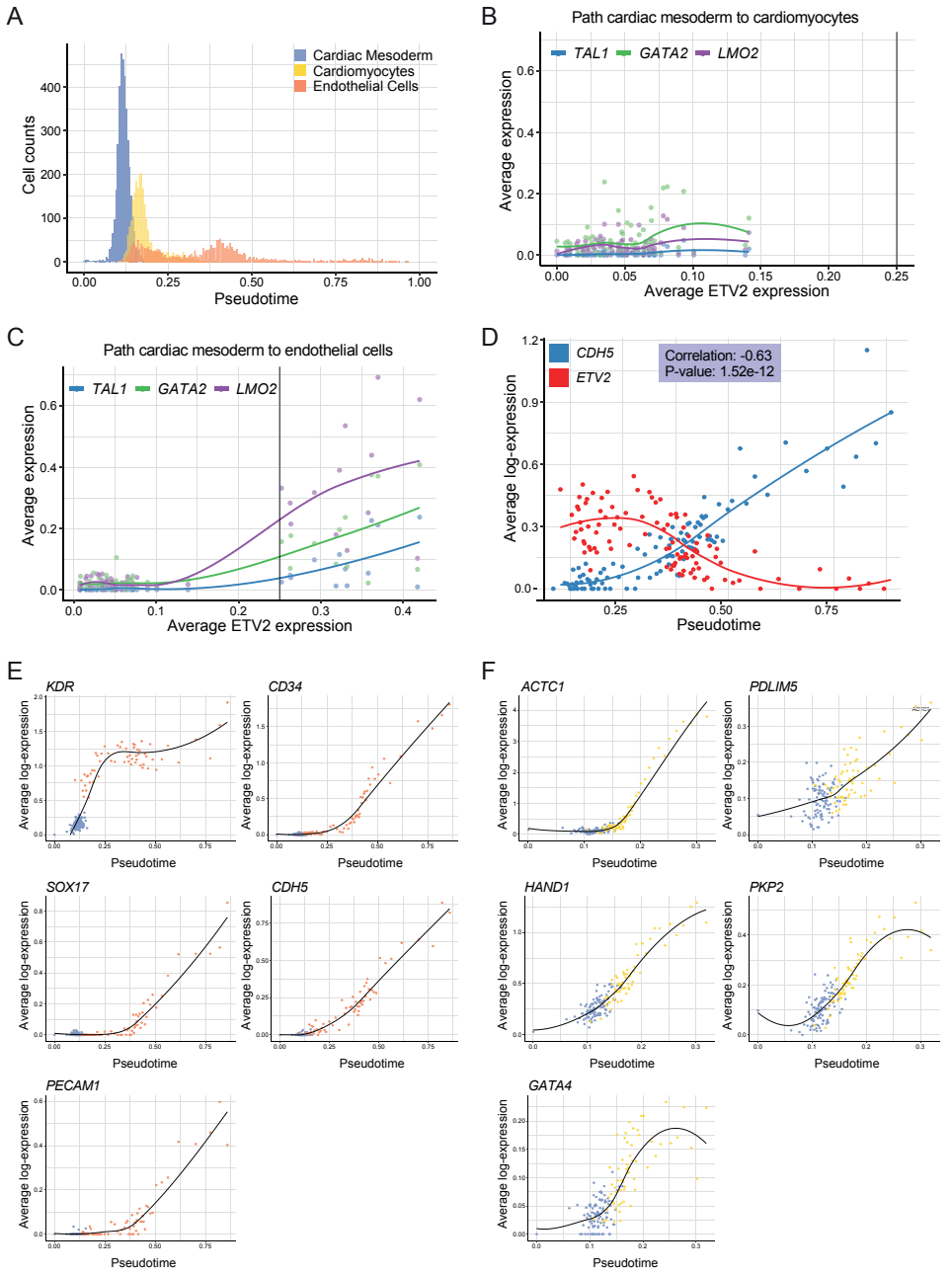


**Figure 3.2: Pseudotime analysis of co-differentiation** (A) Distributions of pseudotime for each cell cluster. (B-C) Average expression of ETV2 and ETV2 target genes along the developmental path of CMs (B) and ECs (C). Threshold (indicated in black) is set to the timepoint where the average ETV2 expression in the endothelial cell cluster reaches 0.25. (D) Average expression of CDH5 and ETV2 in the EC cluster across pseudotime. Binning and averaging was performed as for (B) and (C). The p-value for the correlation between CDH5 and ETV2 expression is based on the null hypothesis that the correlation is zero. (E) Expression of endothelial markers across pseudotime along the development path of ECs. (F) Expression of cardiac markers across pseudotime along the development path of CMs.

Taken together, these analyses confirmed the differentiation of cardiac mesoderm into CMs and ECs, which we had discovered previously. They also revealed the increase of ETV2 as a global indicator of early lineage separation and a transient pulse of high ETV2 at the beginning of EC specification.

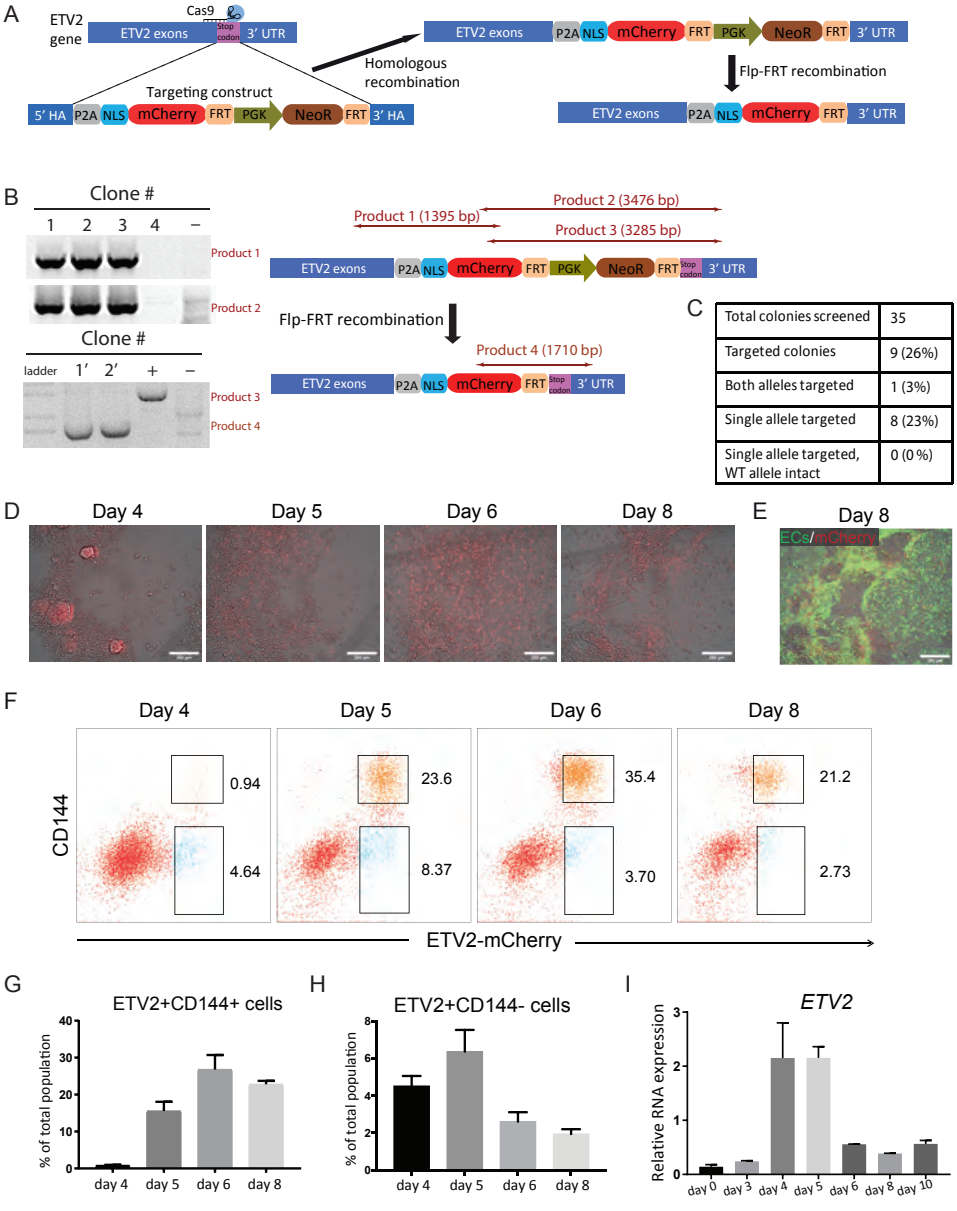


**Figure 3.3: Generation and characterization of an ETV2mCherry hiPSC reporter line. (Caption on the next page.)**

### 3.2.2 Generation and characterization of an ETV2$^{mCherry}$ fluorescent hiPSC reporter line

In order to follow ETV2 expression in real-time, we introduced a fluorescent reporter for ETV2 in hiPSCs. P2A-mCherry with a nuclear localization signal (NLS) and a neomycin selection cassette was inserted into the endogenous ETV2 locus before the stop codon using CRISPR/Cas9-facilitated homologous recombination (Fig. 3.3A). After neomycin selection and excision of the selection cassette, targeted hiPSC clones were validated by PCR (Fig. 3.3B) and Sanger sequencing. Clones with ETV2$^{mCherry}$ in both alleles were selected for downstream analysis (Fig. 3.3C). Time-lapse imaging of differentiating cultures showed the appearance of nuclear mCherry from day 4 of differentiation (Fig. 3.3D and supplemental online Video 1). Co-staining of live cells with EC-specific fluorescent agglutinin showed distinct EC clusters on day 8 of differentiation (Fig. 3.3E). Flow cytometry analysis at different stages of differentiation revealed upregulation of ETV2 (mCherry protein) as early as day 4 of differentiation, followed by the upregulation of the EC-specific marker CD144 (Fig. 3.3F). Quantification of percentages of single positive (SP; ETV2$^{mCherry}$+CD144-)(Fig. 3.3G) and double positive (DP; ETV2$^{mCherry}$+CD144+)(Fig. 3.3H) cells on day 4, 5, 6 and 8 of differentiation from at least three independent experiments showed a decrease and an increase of SP and DP cells respectively (Fig. 3.3G-H). mCherry protein remained present for a longer period than endogenous ETV2 mRNA (Fig. 3.3I), likely due to a longer half-life of the protein compared to the mRNA. This also explains the persistence of mCherry signal in both the DP and SP population.

**Figure 3.3: Generation and characterization of an ETV2mCherry hiPSC reporter line (Figure on the previous page.)** (A) Schematic of CRISPR/Cas9-Mediated Knock-in of mCherry into the ETV2 locus of hiPSCs. mCherry and Neomycin Resistance (NeoR) was inserted into the ETV2 locus through homologous recombination. Then NeoR cassette was removed by flpo recombinase. (B) PCR screening of targeted clones with correct insertion at the ETV2 locus. Two pairs of primers (for product 1 and 2) were used to confirm the integration of the construct. Clone 1, 2, 3 were correctly targeted and clone 4 was not targeted. Non-targeted hiPSCs (-) were included as negative control. Lower panel: The excision of the neomycin-resistance cassette was confirmed by PCR (product 3 and 4 for before and after excision). Clone 1' ad 2' were successfully excised. Genomic DNA before excision (+) and non-targeted hiPSCs (-) were included as positive and negative control, respectively. (C) Summary of CRISPR targeting efficiency. Out of 35 colonies screened, 1 colony was targeted in both alleles. 8 colonies were targeted in 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 differentiation using the ETV2-mCherry reporter line. Scale bar 200 μm. (E) FITC-agglutinin staining of cells at the same location as shown in (D) on day 8. mCherry is in red and Agglutinin is in green. Scale bar 200 μm. (F) Fluorescence activated cell sorting (FACS) based on CD144 and ETV2mCherry expression on day 4, 5, 6 and 8 of differentiation. (G-H) Quantification of ETV2+CD144- (SP) ETV2+CD144+ (DP) cells by flow cytometry on differentiation day 4, 5, 6 and 8. (I) Quantification of ETV2 expression by qPCR on different days of differentiation. (F-G) Error bars are standard deviations calculated from three independent experiments.

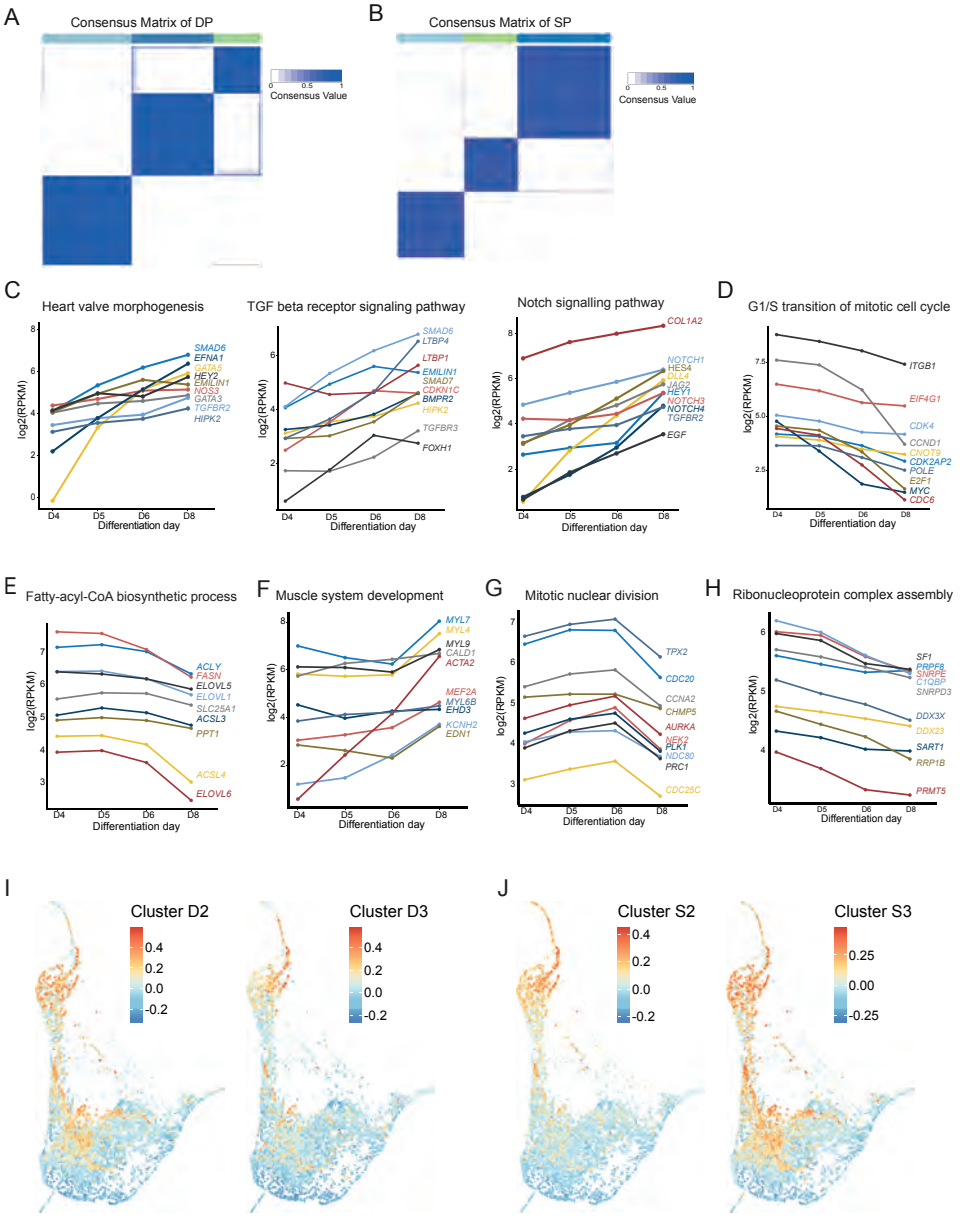


**Figure 3.4: Bulk RNA-seq analysis of of EC and CM differentiation** (A-B) Consensus clustering of the 3000 most variant genes across all DP samples (A) and all SP samples (B). All genes were divided into 3 clusters D1-D3 for DP (A) and S1-S3 for SP (B). Consensus value indicates similarity between two genes. (C-H) Representative GO terms enriched in cluster D1 (C), D2 (D), D3 (E), S1 (F), S2 (G) and S3 (H). Representative genes mapped to these GO terms and their expression levels from day 4 to day 8 are shown. (I-J) Low-dimensional representation of the scRNA-seq data. Each data point is a single cell. Mean expression of genes in cluster D1 (I) and S1 (J) in the scRNA-seq data is indicated by color. Gene expression was scaled gene-wise prior to averaging.

### 3.2.3 The ETV2$^{mCherry}$ fluorescent reporter allows for the purification of differentiating cells with lineage-specific expression profiles



We next sorted DP and SP cells at different stages of differentiation (Fig. 3.3F) and performed bulk RNA-seq of at least three independent replicates. Principal component analysis (PCA) showed that DP and SP populations diverged progressively over time (Fig. 3.5A). Mapping of the bulk transcriptomes to the scRNA-seq data revealed that DP samples aligned on the EC branch and SP cells on the CM branch (Fig. 3.5B). Notably, SP and DP cells of later time points were further away from cardiac mesoderm, reflecting ongoing differentiation (Fig. 3.5B).

We next leveraged the higher sensitivity and accuracy of bulk RNA-seq compared to scRNA-seq, to get a more comprehensive and robust transcriptional characterization of the subpopulations. By consensus clustering of the most variable genes across DP or SP cells (3000 genes each) we found three gene clusters for each population, with distinct expression dynamics (Fig. 3.5C-D, Fig. 3.4A-B, Table S5). In DP cells, cluster D1 (1226 genes) expression increased over time. Gene Ontology (GO) terms enriched in cluster D1 included "angiogenesis", "Notch signaling pathway", "transforming growth factor beta receptor signaling pathway", "receptor-mediated endocytosis" and "developmental maturation" (Fig. 3.5E, Table S6). In accordance with this analysis, angiogenesis-related genes (CDH5, TIE1, TEK, EFNB2, SOX18, VEGFB, LEPR), Notch and transforming growth factor beta receptor signaling 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 (Fig. 3.5G, Fig. 3.4C). In the scRNA-seq data, cluster D1 genes were specifically expressed in the EC cluster and showed increasing expression along pseudotime (Fig. 3.5I). Cluster D1 genes are thus likely involved in EC-specific functions. Cluster D2 (1127 genes), which was downregulated after day 4 (Fig. 3.5C), was enriched for cell cycle-related genes (ITGB1, CDK4, CCND1, CDK2AP2, MYC, CDC6)(3.4D). Cluster D3 (647 genes), which was downregulated after day 5-6 (Fig. 3.5C), contained cell proliferation- and fatty-acyl-CoA biosynthetic process-related genes (ACLY, FASN, ELOVL1, SLC25A1, ACSL3, ACSL4)(Fig. 3.4E). Genes in clusters D2 and D3 were more broadly expressed in the scRNA-seq data (Fig. 3.4I). Their dynamics likely reflect changes in proliferation and metabolism at the exit from the multipotent progenitor state.

In SP cells, cluster S1 (936 genes) increased over time and contained genes enriched for GO terms related to heart development and function (Fig. 3.5F, Table S6). In agreement with this observation, cardiac chamber and cardiac muscle development related genes (MYH6, HAND1, MYH10, TNNT2, NKX2-5, ISL1, TNNC1, MYOD, LMO4 and HEY1, MYL7, MYL4, ACTA2, KCNH2) were upregulated over the course of differentiation (Fig. 3.5H, Fig. 3.4F). Cluster S1 genes were highly expressed in the cardiac mesoderm and CMs clusters in the scRNA-seq data, which showed an increase over pseudotime in the CM lineage (Fig. 3.5J). These genes are thus likely involved in CM-specific functions. Cluster S2 (746 genes), which increased slightly until day 6 and was downregulated afterwards (Fig. 3.5D), contained mitotic nuclear division genes (TPX2, CDC20, NEK2, PLK1, PRC1 and CDC25C) (Fig. 3.4G).

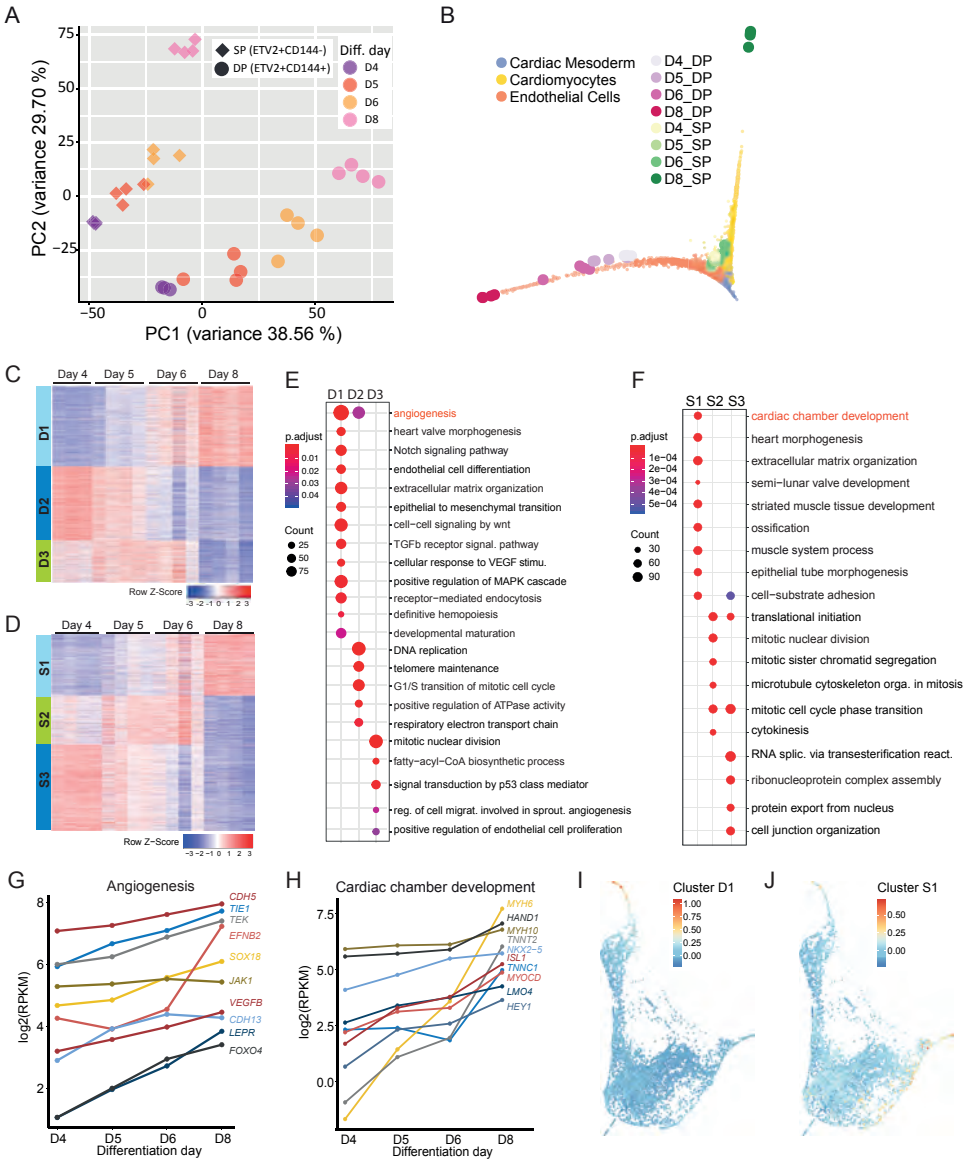


**Figure 3.5: Bulk RNA-seq of the ETV2mCherry reporter line shows diverging transcriptional profiles.** (A) PCA of all sorted DP and SP samples collected from three or four independent differentiations. (B) Low-dimensional representation (diffusion map) of scRNA-seq and bulk RNA-seq samples collected on day 4, 5, 6 and 8. The small data points correspond to individual cells, the large symbols correspond to bulk samples. Different clusters of cells or bulk samples are labelled with different colors. (C-D) Gene expression pattern in all DP (D) and SP (E) cells. The 3000 most variable genes across all DP or SP samples were identified and grouped into three clusters by consensus clustering. The genes in each cluster can be found in Table S5. The color scale represents relative expression (row-wise z-score). (E-F) GO enrichment analysis of each gene cluster of DP (E) and SP (F) samples. Representative GO terms are shown. The complete list of GO terms can be found in Table S6. Color represents the Padjusted of the enrichment analysis and dot size represents the count of genes mapped to the GO term. (G-H) Representative genes mapped to representative GO terms of clusters D1 (G) and S1 (H) and their expression levels from day 4 to day 8 are shown. (I-J) Low-dimensional representation of the scRNA-seq data. Each data point is a single cell. Mean expression of genes in cluster D1 (I) and S1 (J) in the scRNA-seq data is indicated by color. Gene expression was scaled gene-wise prior to averaging.

Cluster S3 (1318 genes), whose expression decreased continuously over time (Fig. 3.5D), contained transcription and translation process-related genes (SF1, SNRPE, DDX23, RRP1B and PRMT5) (3.4H). In the scRNA-seq data, genes from clusters S2 and S3 showed broader expression patterns compared to cluster S1 genes (Fig. 3.4J). The dynamics of clusters S2 and S3 likely reflect changes in proliferation and metabolism in the CM lineage, analogous to the role of clusters D2 and D3 in the EC lineage.
All in all, time-resolved bulk RNA-seq of sorted SP and DP populations confirmed that ETV2-positive cells contained transcriptionally distinct subpopulations. DP cells were part of the EC lineage, while SP cells corresponded to the CM lineage.



### 3.2.4 ETV2+ cells contain lineage-restricted subpopulations

Next, we wanted to find out how far the various subpopulations we identified differed in terms of their further differentiation potential. To that end, we sorted cells on ETV2 reporter levels shortly after the bifurcation (on day 5) and attempted to differentiate them further towards ECs by adding VEGF (Fig. 3.6A-B). After 5 days of additional differentiation, $ETV2^{mCherry}$+ cells produced more than 90% CD144+CD31+ ECs, while $ETV2^{mCherry}$- cells gave rise to only 10-15% ECs (3.6C-D). Only cells derived from $ETV2^{mCherry}$+ cells expressed endothelial-specific markers, as observed by qPCR and immunofluorescence (Fig. 3.6E-H). These cells also upregulated pro-inflammatory markers, such as ICAM-1 and E-Selectin upon TNF-$\alpha$ stimulation (Fig. 3.6I-L), as shown previously for hiPSC-derived ECs [33]. We could thus conclude that the majority of ETV2-positive cells on day 5 have a strong propensity to produce ECs.
Both the analysis of the scRNA-seq data and the time-resolved bulk RNA-seq of sorted cells identified a subpopulation of ETV2+ cells with CM characteristics. We strongly suspected that the differentiation potential of those cells would be restricted to the CM lineage. To test this hypothesis with our reporter line, we co-differentiated cells until day 7. We chose a later time point for this experiment because the majority of cells are past the bifurcation at this point, and it is, therefore, easier to identify the ETV2+ population that does not correspond to early ECs. We co-stained for CD144 and sorted the cells into DP, SP, and double negative (DN) populations. These subpopulations were then further cultured in the presence of VEGF until day 18 (Fig. 3.6A, M). The majority (>80%) of DP cells differentiated into CD144+CD31+ ECs, in agreement with the previous experiment (Fig. 3.6N-O). In contrast, more than 50% of SP and DN cells differentiated into cTnT+ CMs while very few ECs were detected (Fig. 3.6N-O). Interestingly, CMs derived from SP cells seemed to proliferate more and formed a monolayer composed of a contracting cell sheet, while CMs from DN cells proliferated to a lesser extent and produced only a few, separated clusters of contracting cells (supplemental online Video 2). Almost all DP cells on day 18 expressed the EC marker CD31, while only few cells derived from SP and DN cells were positive for CD31 (Fig. 3.6P-R). Most cells derived from SP and DN expressed CM-specific $\alpha$-Actinin and cTnT and showed typical sarcomeric structures (Fig. 3.6Q-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 $\alpha$-Actinin positive CMs were positive for SM22, possibly indicating their immaturity (Fig. 3.6R).

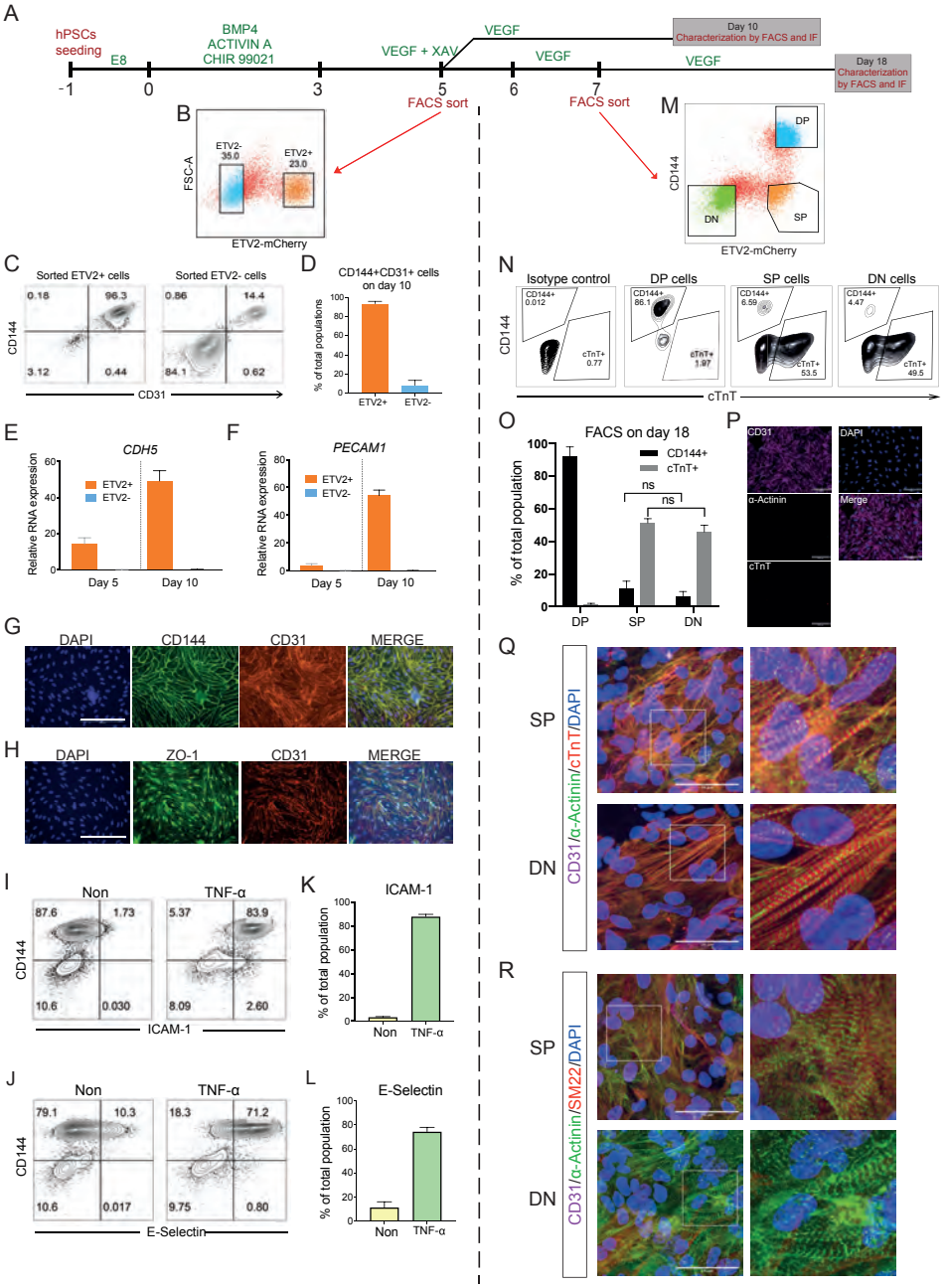


**Figure 3.6: ETV2+ cells contain two lineage-restricted subpopulations (Caption on the next page.)**

Taken together, the VEGF differentiation experiments showed that DP and SP cells are restricted to the EC and CM lineage respectively. DN cells were largely unable to give rise to EC, but produced CMs, albeit with lower efficiency than SP cells. Entering a transient state characterized by high ETV2 expression, thus seems necessary to initiate EC specification.

## 3.3 Discussion

In this study, we characterized the dynamics of EC and CM co-differentiation from hiPSCs [32]. ETV2 was identified as an early indicator of lineage separation and found to be strongly, but transiently, upregulated in ECs, in agreement with its fundamental role in hemangiogenic development [34]. Interestingly, ETV2 expression was also observed in a small population of cardiac mesoderm and CMs. This is reminiscent of a recent study where etv2 expression was observed in lateral plate mesoderm and the CM population in zebrafish [35]. In our experiments, expression of ETV2 target genes seemed to occur only beyond a certain level of ETV2 expression, although this observation could also be explained by a temporal delay between ETV2 upregulation and target gene expression. An ETV2 threshold in our system would be in line with previous reports of an ETV2 threshold in hemangiogenic specification [18, 19]. Our results thus support an ETV2 pulse- and threshold dependent specification of ECs.

With the ETV2$^{mCherry}$ hiPSC reporter line generated to track, isolate and characterize ETV2+ cells, we showed that ETV2+ cells could give rise to both ECs and CMs. Over time, EC and CM precursors acquired more specific endothelial and myocardial identities, respectively, as well as downregulated cell cycle-related genes, which indicated exit from the progenitor state and further maturation.

In the DP subpopulation (EC precursors), several key angiogenesis and Notch signaling pathway genes, like LEPR, FOXO4, DLL4, NOTCH4 and EGF, strongly increased starting from day 4, indicating a specified EC fate but an immature state 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.

**Figure 3.6: ETV2+ cells contain two lineage-restricted subpopulations (Figure on the previous page.)** (A) Schematic of the differentiation protocol and cell sorting. 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 flow cytometry analysis ETV2-mCherry expression on day 5 and gates for cell sorting of ETV2+ and ETV2- cells are shown. (C) Flow cytometry analysis of endothelial markers CD144 and CD31 on day 10 of sorted ETV2+ and ETV2- cell differentiation. (D) Quantification of CD144+CD31+ cells in the total population on day 10 of sorted ETV2+ and ETV2- cell differentiation. (E-F) Quantification of CDH5 and PECAM1 expression 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 for sorted ETV2+ cells on day 10. Scale bar 200 µm. (I-J) Flow cytometry analysis of ICAM1, E-Selectin and CD144 for sorted ETV2+ cells on day 10. Cells were stimulated with TNF-$\alpha$ for 24 h before analysis. (K-L) Quantification of CD144+ICAM-1+ (K) and CD144+E-Selectin+ (L) cells in the population on day 10. (M) Flow cytometry analysis of CD144 and ETV2-mCherry expression on day 7. DP, SP and DN cells were gated and sorted. (N) Flow cytometry analysis of CD144 and CM marker cTnT expression on day 18 of sorted DP, SP and DN cells. Isotype control antibodies were included as negative control. (O) Quantification of CD144+ ECs and cTnT+ CMs on day 18 of DP, SP and DN cell differentiation. (P) Immuno-staining of CD31, a-Actinin, cTnT and DAPI on day 18 of DP cell differentiation. Scale bar 50 µm. (Q-R) 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).

Genes involved in heart development and definitive hematopoiesis were also upregulated during EC development, suggesting cardiac endothelial- and probably a mixture of hemogenic endothelial identity of these ECs. A better characterization of these cells' hematopoietic potential would be interesting, but is beyond the scope of this study. ECs that were further differentiated with VEGF showed a clear endothelial identity and were fully functional based on their inflammatory response upon TNF-$\alpha$ stimulation. Notably, they also expressed a number of cardiac markers like MEOX2, GATA4, GATA6 and ISL1, suggesting a cardiac specific EC identity [32].



The SP subpopulation (CM precursors) had already committed to a cardiac fate on day 4, as evidenced by the expression of cardiac genes HAND1, MYH10, NKX2-5, ISL1, TNNC1, MYOCD and LMO4. However, some crucial genes for CMs were still absent, including MYH6 and TNNT2. MYH6 encodes the major CM thick filament protein MHC-$\alpha$ and TNNT2 is routinely used as a CM marker. Both genes are essential for CM contractility and started to be expressed only after day 4. Their relatively late expression could allow us to identify early and late cardiac progenitors during cardiac development in future studies. CMs were still early progenitors on day 6 of the differentiation as no functionally contracting CMs were observed yet at this stage. Pseudotime analysis also suggested that ECs had differentiated further compared to CMs on day 6. After additional VEGF differentiation, SP cells gave rise to contracting CM, which provided direct evidence they were CM precursors. More importantly, it demonstrated that both ECs and CMs could be derived from ETV2+ progenitors, confirming the presence of a common precursor implied by our earlier studies [32].

Notably, ETV2- cells (DN population) also gave rise to contracting CMs after VEGF treatment, albeit less frequently than SP cells. This difference could be due to either the different cell growth rates or their different developmental origins (FHF vs. SHF). More work is needed to establish the identity of CMs from SP and DN populations in the future.

## 3.4 Conclusion

Transcriptomic analysis at bulk and single cell levels done in this study provide insight into the differentiation dynamics of two important human cardiac lineages and a rich data set for comparison with in vivo data. An ETV2 reporter system was generated in hiPSCs and utilized to identify a new subpopulation of early CM precursors that expressed ETV2.

# 3.5 Materials and Methods

## 3.5.1 Experimental 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 two flippase recognition target (FRT) sites was amplified from the P15 backbone vector (kindly provided by Dr. Konstantinos Anastassiadis, Technical University Dresden). 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 [36]. 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 [32]. 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 [37], 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 Matrigel-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 4℃. 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 CMs 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 S1.

#### 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 S1.

#### 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 S2.

### 3.5.2 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 [38], reads were aligned to the human reference genome GRCh38 using GSNAP (gmap-2014-12-23) [39, 40] and gene read quantification with htseq-count (v0.6.1p1) against the Ensembl v87 annotation [41]. Gene length and GC content bias were normalized using the R package cqn (v1.28.1) [42]. 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) [43]. 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 [44]. 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) [45] and q ≤ 0.05 was considered significant.

### 3.5.3 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

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 (Fig. 3.2A-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 (3.2A-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 [46]. The 5% most highly variable genes (HVGs) 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 [47] and mitochondrial genes. For downstream analysis the top 5% HVGs were used after excluding proliferation [48] and cell cycle [49] related genes.

### Cell cycle analysis

For each combined data set, cell cycle analysis was performed with the scran package using the cyclone function [50] on normalized counts. 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) [51] applied to log-transformed normalized counts. Then, the two replicates were batch corrected with the fast mutual nearest neighbors (MNN) correction method [52] on the cell cycle corrected counts, using the 30 first principal components and 20 nearest-neighbors.



### Clustering

First, 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 = 20). Louvain clustering was applied to the SNN graph using the igraph python package (V 0.7.1) with 0.4 for resolution parameter. This resulted in 5 clusters. Two of these 5 clusters were excluded from further analysis based on the expression of pluripotency markers [53].

### Dimensionality reduction and pseudotime

Dimensionality reduction was performed using the python scanpy pipeline (V 1.4.6). First 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 cellular progression. The graph was projected into two dimensions with the default force-directed graph layout and starting positions obtained from the partition-based graph abstraction (PAGA) output [54]. PAGA estimates connectivities between partitions and performs an improved version of diffusion pseudotime. Diffusion pseudotime [52, 54] was calculated on these graphs with root cells selected based on the graph layout from the “Cardiac Mesoderm” cluster.
Average gene expression trajectories were calculated by dividing the cells of each cluster into bins along pseudotime. 50 bins were created for cardiac mesoderm and 30 bins for each, endothelial cells and cardiomyocytes. Then, the average log-expression per bin was shown. The value of the threshold shown in Fig. 3.1D,E was determined by calculating the point in pseudotime where the average ETV2 expression was the lowest in the endothelial cell cluster before the peak expression, which corresponds to a value around 0.25 [48].

### Differential expression analysis and identification of cluster maker genes

The R package edgeR (V 3.24.3, 31) [43] 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.

#### 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.



### 3.5.4 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/ GEO: GSE157954 (bulk) and GEO: GSE202901 (single cell). Supplementary tables and videos are available at https://doi.org/10.5061/dryad.9p8cz8wkg.

**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). K. Anastassiadis (Technical University Dresden) for providing P15 backbone with a Neomycin resistance cassette surrounded by two FRT sequences and CAGGs-Flpo-IRES-puro vector.

**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). M. M. and S.S. were supported by the Netherlands Organisation for Scientific Research (NWO/OCW,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] W. P. Devine et al. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife*, 3, oct 2014.

[2] F. Lescroart et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature Cell Biology 2014 16:9*, 16(9):829–840, aug 2014.

[3] S. Zaffran et al. Right ventricular myocardium derives from the anterior heart field. *Circulation Research*, 95(3):261–268, aug 2004.

[4] Y. Saga et al. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*, 126(15):3437–3447, aug 1999.

[5] F. Lescroart et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science*, 359(6380):1177–1181, mar 2018.

[6] M. Ema, S. Takahashi, and J. Rossant. Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. *Blood*, 107(1):111–117, jan 2006.

[7] D. J. Garry and E. N. Olson. A Common Progenitor at the Heart of Development. *Cell*, 127(6):1101–1104, dec 2006.

[8] P. P. Tam, M. Parameswaran, S. J. Kinder, and R. P. Weinberger. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. *Development*, 124(9):1631–1642, may 1997.

[9] 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.

[10] S. D. Vincent and M. E. Buckingham. How to Make a Heart: The Origin and Regulation of Cardiac Progenitor Cells. *Current Topics in Developmental Biology*, 90(C):1–41, jan 2010.

[11] D. Galli et al. Atrial myocardium derives from the posterior region of the second heart field, which acquires left-right identity as Pitx2c is expressed. *Development*, 135(6):1157–1167, mar 2008.

[12] C. L. Cai et al. Isl1 Identifies a Cardiac Progenitor Population that Proliferates Prior to Differentiation and Contributes a Majority of Cells to the Heart. *Developmental Cell*, 5(6):877–889, dec 2003.

[13] N. Paffett-Lugassy et al. Heart field origin of great vessel precursors relies on nkx2.5-mediated vasculogenesis. *Nature Cell Biology 2013 15:11*, 15(11):1362–1369, oct 2013.

[14] A. Ferdous et al. Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo. *Proceedings of the National Academy of Sciences of the United States of America*, 106(3):814–819, jan 2009.

[15] S. De Val et al. Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell*, 135(6):1053, dec 2008.

[16] D. Lee et al. ER71 Acts Downstream of BMP, Notch, and Wnt Signaling in Blood and Vessel Progenitor Specification. *Cell Stem Cell*, 2(5):497–507, may 2008.

[17] F. Liu et al. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO reports*, 16(5):654–669, may 2015.

[18] H. Zhao and K. Choi. A CRISPR screen identifies genes controlling Etv2 threshold expression in murine hemangiogenic fate commitment. *Nature Communications 2017 8:1*, 8(1):1–12, sep 2017.

[19] H. Zhao and K. Choi. Single cell transcriptome dynamics from pluripotency to FLK1+ mesoderm. *Development (Cambridge)*, 146(23), dec 2019.

[20] R. Morita et al. ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, 112(1):160–165, jan 2015.

[21] I. Elcheva et al. Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. *Nature Communications 2014 5:1*, 5(1):1–11, jul 2014.

[22] A. G. Lindgren, M. B. Veldman, and S. Lin. ETV2 expression increases the efficiency of primitive endothelial cell derivation from human embryonic stem cells. *Cell Regeneration*, 4(1):4:1, jan 2015.

[23] V. S. Brok-Volchanskaya et al. Effective and Rapid Generation of Functional Neutrophils from Induced Pluripotent Stem Cells Using ETV2-Modified mRNA. *Stem Cell Reports*, 13(6):1099–1110, dec 2019.

[24] K. Suknuntha et al. Optimization of Synthetic mRNA for Highly Efficient Translation and its Application in the Generation of Endothelial and Hematopoietic Cells from Human and Primate Pluripotent Stem Cells. *Stem Cell Reviews and Reports*, 14(4):525–534, aug 2018.

[25] B. Cakir et al. Engineering of human brain organoids with a functional vascular-like system. *Nature Methods 2019 16:11*, 16(11):1169–1175, oct 2019.

[26] K. Wang et al. Robust differentiation of human pluripotent stem cells into endothelial cells via temporal modulation of ETV2 with modified mRNA. *Science Advances*, 6(30), jul 2020.

[27] B. Palikuqi et al. Adaptable haemodynamic endothelial cells for organogenesis and tumorigenesis. *Nature 2020 585:7825*, 585(7825):426–432, sep 2020.

[28] D. T. Paik et al. Large-scale single-cell RNA-seq reveals molecular signatures of heterogeneous populations of human induced pluripotent stem cell-derived endothelial cells. *Circulation Research*, 123(4):443–450, 2018.

[29] I. R. McCracken et al. Transcriptional dynamics of pluripotent stem cell-derived endothelial cell differentiation revealed by single-cell RNA sequencing. *European Heart Journal*, 41(9):1024–1036, mar 2020.

[30] S. C. Den Hartogh et al. Dual reporter MESP1 mCherry/w-NKX2-5 eGFP/w hESCs enable studying early human cardiac differentiation. *Stem Cells (Dayton, Ohio)*, 33(1):56–67, jan 2015.

[31] A. Moretti et al. Multipotent Embryonic Isl1+ Progenitor Cells Lead to Cardiac, Smooth Muscle, and Endothelial Cell Diversification. *Cell*, 127(6):1151–1165, dec 2006.

[32] 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.

[33] O. V. Halaidych et al. Inflammatory Responses and Barrier Function of Endothelial Cells Derived from Human Induced Pluripotent Stem Cells. *Stem Cell Reports*, 10(5):1642–1656, may 2018.

[34] N. Koyano-Nakagawa and D. J. Garry. Etv2 as an essential regulator of mesodermal lineage development. *Cardiovascular Research*, 113(11):1294–1306, sep 2017.

[35] B. Chestnut, S. Casie Chetty, A. L. Koenig, and S. Sumanas. Single-cell transcriptomic analysis identifies the conversion of zebrafish Etv2-deficient vascular progenitors into skeletal muscle. *Nature Communications 2020 11:1*, 11(1):1–16, jun 2020.

[36] A. Kranz et al. An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. *genesis*, 48(8):512–520, aug 2010.

[37] 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.

[38] M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12, may 2011.

[39] 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.

[40] 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.

[41] A. Yates et al. Ensembl 2016. *Nucleic Acids Research*, 44(D1):D710–D716, jan 2016.

[42] 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.

[43] 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.

[44] T. Xu et al. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics (Oxford, England)*, 33(19):3131–3133, oct 2017.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] A. Scialdone et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods*, 85:54–61, sep 2015.

[51] 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.

[52] L. Haghverdi et al. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods 2016 13:10*, 13(10):845–848, aug 2016.

[53] 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.

[54] 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.