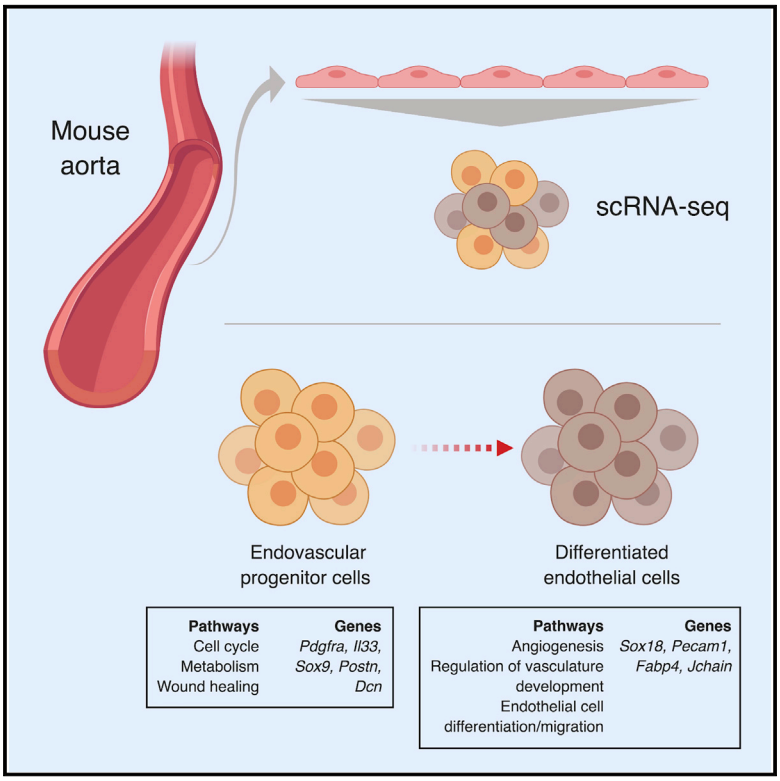
Article

Cell Reports

Single-Cell Transcriptional Profiling of Aortic Endothelium Identifies a Hierarchy from Endovascular Progenitors to Differentiated Cells

Graphical Abstract



Highlights

- scRNA sequencing reveals two distinct aortic endothelial populations
- Sequencing analysis confirms transitionary cells rather than discrete populations
- Endovascular progenitors are slow cycling and have distinct mitochondrial content

Authors

Samuel W. Lukowski, Jatin Patel, Stacey B. Andersen, ..., Ingrid Winkler, Joseph E. Powell, Kiarash Khosrotehrani

Correspondence

k.khosrotehrani@uq.edu.au

In Brief

Lukowski et al. demonstrate the existence of two distinct endothelial cell populations in the aorta via single-cell sequencing. The data confirm that a progenitor population transitions to a mature endothelial cell, defining an endothelial hierarchy.

Lukowski et al., 2019, Cell Reports *27*, 2748–2758
May 28, 2019
https://doi.org/10.1016/j.celrep.2019.04.102

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

Single-Cell Transcriptional Profiling of Aortic Endothelium Identifies a Hierarchy from Endovascular Progenitors to Differentiated Cells

Samuel W. Lukowski,[1,6] Jatin Patel,[2,6] Stacey B. Andersen,[1] Seen-Ling Sim,[2] Ho Yi Wong,[2] Joshua Tay,[5] Ingrid Winkler,[5] Joseph E. Powell,[1,3,4,7] and Kiarash Khosrotehrani[2,7,8,*]

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia
[2]The University of Queensland Diamantina Institute, Brisbane, QLD 4102, Australia
[3]Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[4]St. Vincent's Clinical School, University of New South Wales Sydney, Sydney, NSW 2010, Australia
[5]Faculty of Medicine, Translational Research Institute, Mater Research Institute-The University of Queensland, Woolloongabba, QLD 4102, Australia
[6]These authors contributed equally
[7]Senior author
[8]Lead Contact
*Correspondence: k.khosrotehrani@uq.edu.au
https://doi.org/10.1016/j.celrep.2019.04.102

SUMMARY

The cellular and molecular profiles that govern the endothelial heterogeneity of the circulatory system have yet to be elucidated. Using a data-driven approach to study the endothelial compartment via single-cell RNA sequencing, we characterized cell subpopulations within and assigned them to a defined endothelial hierarchy. We show that two transcriptionally distinct endothelial populations exist within the aorta and, using two independent trajectory analysis methods, confirm that they represent transitioning cells rather than discrete cell types. Gene co-expression analysis revealed crucial regulatory networks underlying each population, including significant metabolic gene networks in progenitor cells. Using mitochondrial activity assays and phenotyping, we confirm that endovascular progenitors display higher mitochondrial content compared to differentiated endothelial cells. The identities of these populations were further validated against bulk RNA sequencing (RNA-seq) data obtained from normal and tumor-derived vasculature. Our findings validate the heterogeneity of the aortic endothelium and previously suggested hierarchy between progenitor and differentiated cells.

INTRODUCTION

The circulatory system throughout an organism is lined by specialized endothelial cells that form the lumen of the vasculature tube (Herbert and Stainier, 2011). It is becoming apparent that, within each organ bed, the endothelial compartment is heterogeneous. This includes variations between vessel types (arterial, venous, and lymphatic), location (macro- or microvascular), and between neighboring endothelial cells (Adams and Alitalo, 2007; Eichmann et al., 2005), with multiple subpopulations being found, such as tip and stalk cells (Iruela-Arispe and Davis, 2009). Although each of these situations has been examined and validated individually on a candidate-based approach, there has never been an unbiased attempt to describe the heterogeneity that characterizes this compartment (Aird, 2012). Furthermore, although each aspect of heterogeneity has been explored individually, the overlap between the various indicators of heterogeneity is currently unknown. Importantly, the complexity in vessel regeneration or turnover compounds the need to better understand the interactions between the endothelial populations and ascertain the specific genes that control these processes (Aird, 2007a, 2007b).

We recently reported that in a variety of vascular beds reside an endothelial stem cell population that was termed endovascular progenitor (EVP) (Patel et al., 2017). These cells are governed by a stringent set of functional characteristics, such as slow cycling, self-renewal, engraftment potential, and finally differentiation to transit amplifying and mature endothelial cells. In the initial description, making use of common endothelial markers, these EVP cells had a stringent cell surface marker expression profile, $CD34^{+}CD45^{-}CD31^{lo}VEGFR2^{-}$, and also gave rise to a transit amplifying (TA) population, which then transitioned to a definitive differentiated (D) endothelial cell with increasing surface levels of VEGFR2 and CD31. This endothelial hierarchy was observed not only in situations of homeostasis (aorta) but also of pathology, such as wound healing or cancer. Bulk RNA sequencing was conducted on EVP and D cells from the aorta, demonstrating remarkable gene expression changes during the transition. However, others have described different progenitor populations with endothelial potential characterized by the expression of different genes, such as CD157, c-Kit, or activated protein C (Fang et al., 2012; Wakabayashi et al., 2018; Yu et al., 2016). It is therefore difficult to establish the relationship between these different cell populations emanating from different studies.

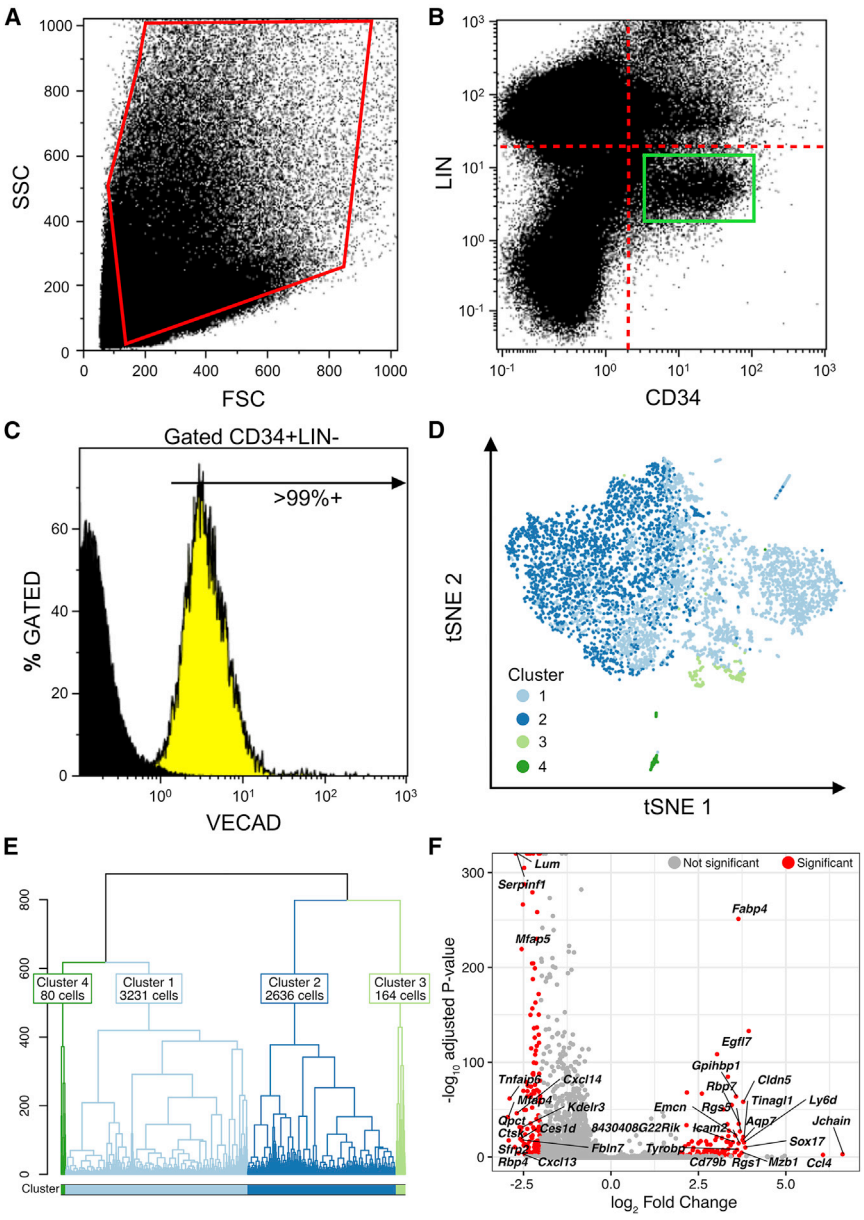


Figure 1. scRNA-Seq and FACS of Aortic Endothelium

(A and B) Flow cytometry plots demonstrating isolation of aortic endothelial cells to exclude debris and doublets (A) and that that are Lineage-CD34+ (B).

(C) >99% of Lineage-CD34+ cells are VE-cadherin+ (VECAD, yellow; negative population, black).

(D) A t-SNE plot of 6,111 cells colored by cluster assignment. Clusters were determined using unbiased hierarchical clustering.

(E) The dendrogram illustrates the structure of the clustered cells.

(F) Volcano plot of genes that differentially expressed between clusters 1 and 2. The red dots represent significant genes determined as having an absolute fold change greater than 2 and are below the false discovery rate threshold of 0.01 (1%). Black dots are genes that are not significant. The x axis shows the $\log_2$ fold change of each gene, and the y axis shows the −log10 false discovery rate (FDR)-adjusted p value.

RESULTS

scRNA-Seq

We obtained an unbiased study of the aorta endothelium by dissecting the descending aorta of 6- to 8-week-old mice and proceeded to cell separation of the endothelium, based on living $Lin^{-}CD34^{+}$ cells (Figures 1A and 1B). We further demonstrate that >99% of this population of cells express VE-cadherin (VECAD) to highlight their restricted endothelial lineage (Figure 1C). Sequence data from three scRNA-seq libraries constructed from three pools of 30 mouse aorta samples were combined for processing and analysis. Pre-quality control (QC) libraries contained 2,212, 3,256, and 2,203 cells, respectively, a total of 7,671 cells. We detected a median of 1,850 unique molecular identifiers (UMIs) per cell, a median of 788 genes per cell, and obtained an average read depth of 53,631 reads per cell. This sequencing depth allowed us to reach saturation for the number of detected genes per cell in each of the 3 samples (Figure S1). Following quality control processing in the *ascend* package, we obtained 1,638, 2,821, and 1,652 cells for a total of 6,111 high-quality cells in the final dataset.

The recent development of single-cell RNA sequencing (scRNA-seq), which combines single-cell isolation techniques with RNA-seq, creates an opportunity to study the transcriptomes of individual isolated cells. This enables clear distinction to be made between subpopulations and the thorough assessment and quantification of gene transcripts in an unbiased way without preconception. In this study, we aimed at characterizing the gene expression in the endothelium of the aorta at single-cell resolution to obtain a clear and unbiased picture of endothelial cell populations. Our findings largely confirm the previously described hierarchy between a slow-cycling EVP and a terminally differentiated D population, opening an understanding of the homeostasis of the vasculature.

Identification of Distinct Endothelial Clusters in Aorta Endothelium

Unsupervised hierarchical clustering analysis of the scRNA-seq data using the Clustering at Optimal Resolution (CORE) method (Nguyen et al., 2018) revealed four stable cell populations containing 3,231 (52.87%), 2,636 (43.14%), 164 (2.68%), and 80 (1.31%) cells, respectively. These clusters were stable across

Table 1. Top 20 Differentially Expressed Genes per Cluster (Cluster *x* versus Remaining Clusters) Sorted by Highest Mean Expression and Log_2 Fold Change

C1 Gene	C1 log_2FC	C2 Gene	C2 log_2FC	C3 Gene	C3 log_2FC	C4 Gene	C4 log_2FC
Jchain	6.61	*Serpinf1*	2.78	*H2-Ab1*	6.16	*Tnnc2*	10.33
Cldn5	3.85	*Lum*	2.78	*Hmgb2*	6.08	*Acta1*	10.28
Rgs5	3.76	*Mfap5*	2.62	*H2-Eb1*	5.89	*Sln*	10.24
Egfl7	3.73	*Smoc2*	2.59	*H2-Aa*	5.87	*Myl7*	10.15
Fabp4	3.66	*Pcolce*	2.54	*Cd74*	5.36	*Myl1*	9.98
Ccl5	3.58	*Gsn*	2.47	*H2afz*	4.86	*Myl4*	9.88
Rbp7	3.57	*Ccl11*	2.47	*Hmgb1*	3.43	*Tnnt2*	9.82
Gpihbp1	3.55	*Dcn*	2.47	*Ptma*	3.38	*Ckm*	9.81
Esam	3.31	*Clec3b*	2.46	*Tmsb4x*	3.00	*Mb*	9.73
Ccl4	3.25	*Nupr1*	2.36	*Cst3*	2.93	*Tnni3*	9.73
Hbb-bt	3.03	*Pi16*	2.36	*Pfn1*	2.87	*Mylpf*	9.18
Cd36	2.93	*Col1a2*	2.36	*Actb*	2.80	*Actc1*	9.18
Hba-a2	2.88	*Col3a1*	2.30	*Actg1*	2.68	*Fabp3*	9.13
Hba-a1	2.85	*Serping1*	2.29	*Ppia*	2.62	*Cox6a2*	8.86
Hbb-bs	2.58	*Igfbp6*	2.23	*Gm9843*	2.47	*Tpm1*	8.16
Rgcc	2.37	*Timp2*	2.17	*Fau*	2.38	*Slc25a4*	5.22
Tspan13	2.33	*Rarres2*	2.14	*Crip1*	2.20	*Atp5g1*	4.35
Id1	2.26	*Serpinh1*	2.13	*Eef1a1*	2.16	*Atp5e*	3.04
Aqp1	2.21	*Cd63*	2.10	*Tmsb10*	2.08	*Cox7c*	2.87
Cytl1	2.17	*Gpx3*	2.01	*Oaz1*	2.00	*Cox6c*	2.84

FC, fold change

the three batch experiments, giving confidence that they represented the biology of the homeostatic aortic endothelium (Figures 1D and 1E). To identify the cell types represented by each cluster, we performed differential gene expression analysis, comparing the cells in each cluster against the remaining cells. Our results revealed that cells in cluster one displayed a mature, differentiated vascular endothelial signature with increased expression of specific genes, including *Sox17*, *Cd36*, *Fabp4*, and *Cldn5* (Figure 1F). On the other hand, cells in cluster two displayed a more mesenchymal phenotype and overexpressed genes, such as *Dcn*, *Lum*, and *Gsn* (Patel et al., 2017). Cells belonging to cluster three also displayed a distinct terminally differentiated endothelial phenotype but also expressed markers of inflammation, including *Il1b* and *Aif1*, compared to other endothelial cells in clusters one and two. Cells in cluster four were identified as either cardiac or skeletal muscle due to their high expression of troponin and other myocyte markers, including *Tnnt2*, *Tnni3*, and *Myl7* (cardiac) and *Tnnc2*, *Tnnt3*, *Myl1*, and *Myl4* (striated muscle). Although of interest, this fourth cluster represented only a small fraction of the entire sample, and its relevance to endothelial biology was uncertain. Hence, we proceeded to further analysis by considering this fourth cluster as a contaminant. Table 1 shows the top differentially expressed genes per cluster sorted by mean expression and log_2 fold change. The complete results are presented in Table S1.

Expression Differences between EVP and D Populations

Previously, we detected EVP progenitors and D differentiated cells in aortic endothelium and tumor tissue using whole-tissue RNA-seq (Donovan et al., 2019; Patel et al., 2017). These populations were shown to express a set of genes that place cells along the continuum of differentiation (*Sox9*, *Il33*, and *Pdgfra* in EVP and *Pecam1* and *Sox18* in D). To further investigate the transcriptional differences between populations at single-cell resolution, we restricted our analysis to clusters one and two. A differential expression analysis between these subpopulations revealed 234 genes (log_2 fold change > 2; adj. p value < 0.01; Figure 1D). Of these, 90 were upregulated in cluster one and 144 were upregulated in cluster two. We observed cells in cluster one expressed *Pecam1* (log_2 fold change = 3.06; adj. $p = 2.53 \times 10^{-22}$) and *Cdh5* (log_2 fold change = 2.91; adj. $p = 4.05 \times 10^{-14}$) at significantly higher levels compared to cells in cluster two as observed in the D population. Conversely, cells in cluster two showed increased expression of *Pdgfra* (log_2 fold change = 2.00; adj. $p = 1.02 \times 10^{-81}$), *Il33* (log_2 fold change = 2.06; adj. $p = 3.46 \times 10^{-15}$), and *Sox9* (log_2 fold change = 2.18; adj. $p = 8.94 \times 10^{-3}$), as observed in EVP (Patel et al., 2017).

Using these marker genes, we visualized cells belonging to clusters one and two using t-Distributed Stochastic Neighbor (t-SNE) dimensionality reduction and observed a clear separation between cells expressing *Cdh5* and *Pecam1* (D; cluster one) and those expressing *Sox9*, *Il33*, and *Pdgfra* (EVP; cluster two; Figure 2A). These findings indicate that clusters one and two correspond to transcriptionally distinct populations of endothelial cells at different stages of differentiation.

Marker genes derived from the literature are used to establish cell identity and infer function of clustered cells; however, many canonical marker genes do not clearly show subtle changes in

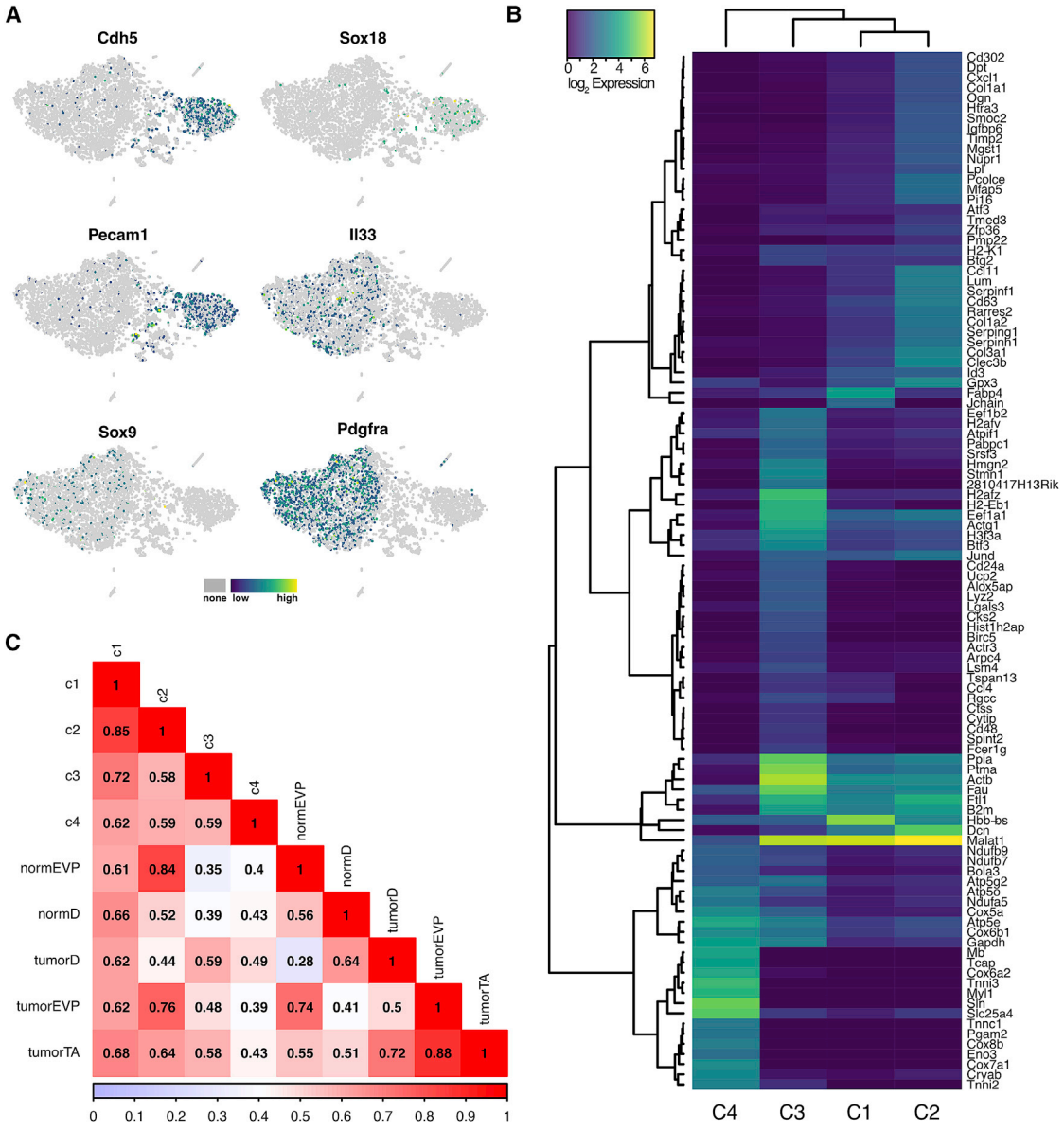


Figure 2. Expression of Marker Genes Defining Each Population

(A) Expression levels of six canonical marker genes were overlaid on the 2D t-SNE plot and highlight the cell specificity of their expression. The expression levels are shown as $\log_2$(counts + 1)-transformed values, and the gradient represents low (gray) to high values (purple).

(B) Heatmap of LASSO-identified marker genes. Marker genes identified using the LASSO machine learning approach clearly mark each cluster and explain the most deviance per cluster. Expression levels are plotted as $\log_2$(counts + 1)-transformed values. The color gradient shows low (purple) to high expression levels (yellow).

(C) Pairwise Spearman's correlations were performed between single-cell data for each cluster and bulk RNA-sequencing data for normal or tumor populations of D and EVP. The correlation coefficient for each test is shown in the center of each square, and the color gradient shows the range of correlation coefficients from 0 (blue) to 1 (red).

gene expression that exist between closely related cell populations. Machine learning approaches have been used to refine gene lists that capture the transcriptional signature of cell populations in scRNA-seq analyses in order to produce a list of genes that explain the greatest amount of deviance (Nguyen et al., 2018). Using a least absolute shrinkage and selection operator (LASSO) regression approach, we further characterized each cell cluster and established a set of gene markers describing the transcriptional differences between each of the four clusters. Starting with differentially expressed genes obtained for each individual cluster, we used the LASSO regression procedure to detect gene markers for each of the four clusters, obtaining 162, 135, 33, and 25 genes for clusters 1–4, respectively. Our results show that each cluster expresses a subset of genes

distinguishing it from the remaining clusters. For example, cluster one (D cells) expressed higher levels of fatty acid binding protein 4 (*Fabp4*) and *Jchain*, and cluster two (EVP) showed increased expression of several serpins (*Serpinf1*, *Serping1*, and *Serpinh1*), *Ccl11*, *Postn*, and *Dcn*. Cluster three had markedly higher expression of important transcription or translation factors, including *H2afz*, *Btf3*, *H3f3a*, and *Eef1a1*, as well as *Ppia*, which encodes cyclophilin A. Finally, cluster four cells expressed genes related to cardiac muscle, including *Mb*, *Tnni3*, and *Tnnc1*, which confirmed our initial findings. Expression levels of across all clusters are shown for 103 genes in Figure 2B, and the table of genes per cluster and their associated deviance coefficients are presented in Table S1.

Correlation of Whole-Tissue D and EVP to scRNA-Seq Clusters One and Two

The single-cell transcriptional profiles that we generated allow a high-resolution investigation of endothelial subpopulations. Our previous study of EVP and D populations in aortic endothelium used a bulk RNA-seq approach to examine gene expression in whole-tissue samples (Patel et al., 2017). To confirm that the cell populations in single-cell clusters one and two were consistent with the whole-tissue data D and EVP, respectively, we performed pairwise Spearman's correlations between single-cell clusters and bulk transcriptome data. The gene expression profiles of each were compared using the significantly differentially expressed genes marking each of the four major scRNA-seq populations or the whole-tissue EVP and D populations, a total of 2,105 genes. We observed a very strong correlation for cluster two (proposed EVP population) with the whole-tissue EVP (Spearman's rho = 0.83; $p < 2.2 \times 10^{-16}$), and the cluster one (proposed D population) cells were well correlated with whole tissue D (Spearman's rho = 0.64; $p = 1.28 \times 10^{-239}$). We also observed a moderate, albeit less significant, correlation of whole-tissue EVP with cluster one cells (Spearman's rho = 0.60; $p = 9.03 \times 10^{-202}$; Figure 2C).

We also tested bulk RNA-seq data that we generated for EVP and D endothelial populations isolated from a B16 model of melanoma (Donovan et al., 2019). In these tumor vasculature populations, we also observed a strong correlation of tumor EVP with normal tissue EVP (Spearman's rho = 0.75; $p < 2.2 \times 10^{-16}$) and cluster two (Spearman's rho = 0.78; $p < 2.2 \times 10^{-16}$), and the tumor D population correlated well with normal tissue D (Spearman's rho = 0.61; $p < 2.2 \times 10^{-16}$) and cluster one (Spearman's rho = 0.58; $p = 4.28 \times 10^{-190}$). We also noted the tumor EVP population correlated with cluster one and may reflect higher heterogeneity and/or deranged gene expression in the tumor. Of interest, cluster three best correlated with tumor D populations, suggesting that it consists of a population of differentiated cells (D′) that can be observed in situations of injury.

Together, the results of the pairwise correlation analysis further support the classification of cluster one as the D endothelial cell population and cluster two as the EVP population.

Linear Modeling of Gene Expression in D versus EVP Populations

To identify further biologically important differences between the D and EVP populations, we applied a linear modeling approach to investigate changes in the expression levels of each non-zero expressed gene (n = 17,175) per individual cell (m = 5,867) against their assigned cluster, either cluster one (D) or cluster two (EVP). A Bonferroni correction for multiple testing was used to determine statistical significance per gene ($p = 2.89 \times 10^{-6}$), and from these, we retained 4,616 significant results. Our data revealed the top genes with significant changes in expression between D and EVP populations, including *Gsn* (beta = 52.5; $p < 2.2 \times 10^{-16}$), *Dcn* (beta = 25.6; $p < 2.2 \times 10^{-16}$), *S100a6* (beta = 10.6; $p < 2.2 \times 10^{-16}$), *Ftl1* (beta = 10.1; $p < 2.2 \times 10^{-16}$), *Fth1* (beta = 11.0; $p < 2.2 \times 10^{-16}$), *Fabp4* (beta = −11.8; $p = 3.7 \times 10^{-153}$), *Hbb-bs* (beta = −37.1; $p = 2.4 \times 10^{-6}$), *Hba-a1* (beta = −27.31; $p = 6.1 \times 10^{-7}$), and *Hba-a2* (beta = −23.7; $p = 7.3 \times 10^{-7}$). Overall, we detected 4,167 genes with significantly increased expression in EVP cells (mean beta = 0.23) and 449 with increased expression in D cells (mean beta = −0.34). We noted that several of these genes were also detected as innovative marker genes using the aforementioned LASSO method. Here, a positive beta value indicates that, for each unit change in gene expression in D, expression was increased in EVP by the beta value and vice versa. The results of the linear modeling analysis are presented in Table S1.

Co-expressed genes in the D or EVP populations are likely to form regulatory networks specific to the cell type. Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) was performed using all 449 upregulated D genes and the top 449 (out of 4,167 total) EVP genes, and two robust modules were detected in each cell population (Figures 3A and 3B). The D modules contained 194 (blue) and 139 (orange) genes (Figure 3A), and EVP modules contained 121 (blue) and 257 (orange; Figure 3B). We tested whether the genes in each module share important biological functions and observed significant Gene Ontology (GO) term enrichment for cell-type-specific biological processes. In D cells, we identified significant genes underlying endothelial cell differentiation and vasculogenesis (Figures 3C and 3E). The genes in the EVP modules were enriched for cell metabolism, extracellular matrix organization, and epithelial cell proliferation (Figures 3D and 3F). StringDB (Szklarczyk et al., 2015) analysis further confirmed the enrichment of functions tightly linked to cell adhesion, cell differentiation, and endothelium development (D) and metabolism and vasculature development (EVP; Figures S2, S3, S4, and S5).

Using the top 449 genes from the linear regression analysis with negative (D) or positive (EVP) beta values, we interrogated the Enrichr database (Chen et al., 2013) to determine whether those genes increased in D or in EVP represented known cell or tissue types. For genes upregulated in D cells, we observed a highly significant overlap with human umbilical vein endothelial cells (HUVECs) (Z score = −1.55; adj. $p = 4 \times 10^{-41}$), and for genes upregulated in EVP cells, we observed a similar but less significant overlap reflecting their endothelial origin (Z score = −1.44; adj. $p = 1.8 \times 10^{-12}$). Notably, for EVP cells, we observed an epithelial-fibroblast phenotype, such as normal human dermal fibroblast cells (Z score = −1.68; adj. $p = 1.76 \times 10^{-41}$) and normal human bronchial epithelial cells (Z score = −1.64; adj. $p = 1.34 \times 10^{-13}$). The expression of low levels of hemoglobin genes in most endothelial cells might be related to the sensitivity of the method and can also be identified in previous

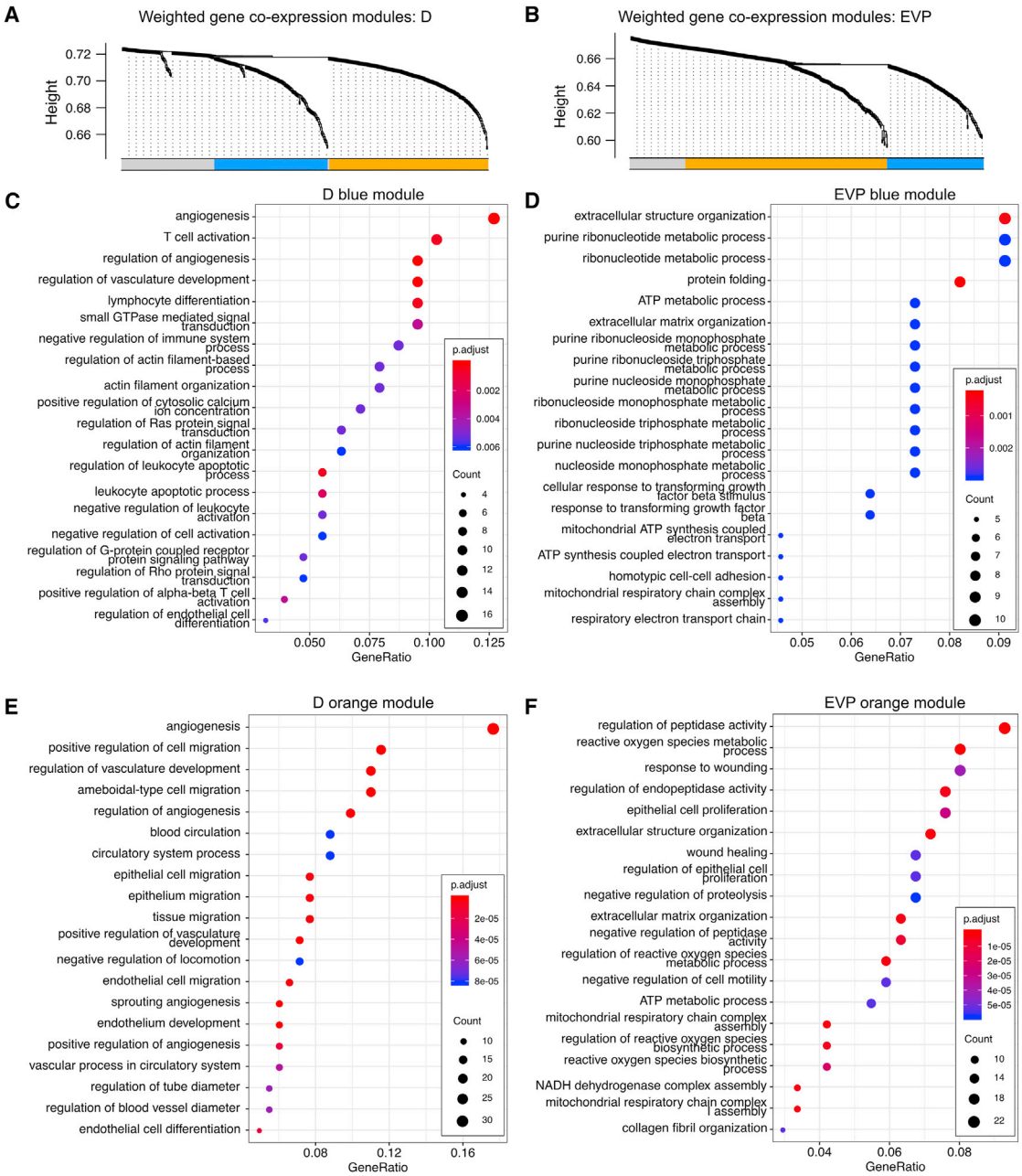


Figure 3. Gene Co-expression Networks and Correlation Analysis

(A and B) Weighted gene co-expression analysis of D (A) and EVP (B) populations revealed two regulatory modules in each, which are shown as orange and blue. (C–F) The genes from each module were analyzed for significantly enriched biological processes using gene ontology annotations for the D population (C and E) and the EVP population (D and F). The size of the circles in (C)–(F) represents the number of significantly enriched genes per module in each process. The adjusted p value (false discovery rate) threshold was 0.05, and significance is shown as a gradient of color in each circle from blue (less significant) to red (more significant).

endothelial-related scRNA datasets (Emara et al., 2014; Newton et al., 2006; Saha et al., 2014, 2017; Zhao et al., 2018).

Validation of Subpopulation Clusters One and Two

Flow cytometry analysis was conducted to demonstrate that cluster one (D) and two (EVP) subpopulations fit our proposed endothelial hierarchy based on CD31 and VEGFR2 levels (Figure 4A). To validate the gene expression findings detailed earlier, we fluorescence-activated cell sorting (FACS)-sorted EVP and D cells from the aorta as demonstrated in Figure 4A. Through immunofluorescence, we show that markers, such as decorin (*Dcn*) and periostin (*Postn*), had higher expression in EVP cells compared to D at the protein level, as suggested by the scRNA-seq data (Figure 4B).

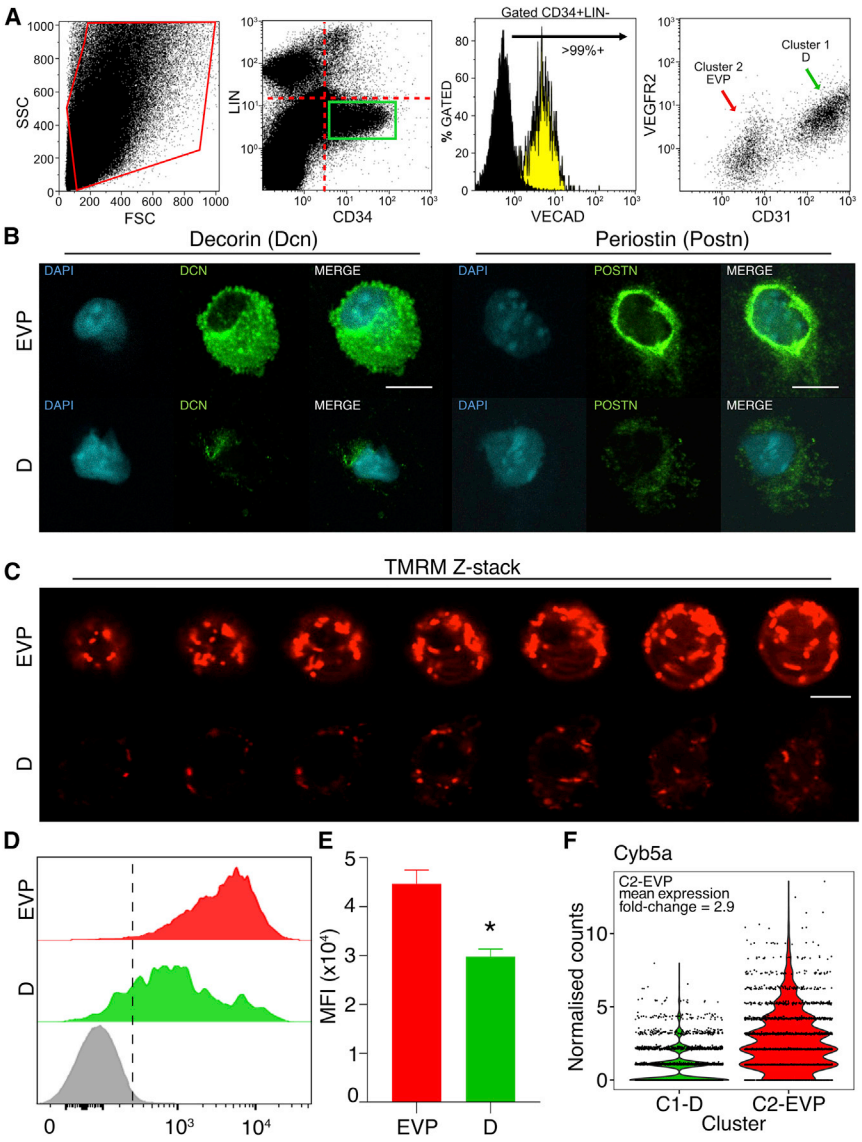


Figure 4. Validation of Marker Genes and Metabolic Activity

(A) Flow cytometry plots to validate that the clusters fall within the predetermined endothelial hierarchy, based on CD31 and VEGFR2 expression (D, definitive differentiated endothelial cell; EVP, endovascular progenitor).

(B and C) Isolated EVP and D cells from aorta were stained for decorin (Dcn) and periostin (Postn) (B) and mitochondrial activity (TMRM; C), with higher expression observed in EVP for each marker under immunofluorescence conditions (n = 3; scale bars represent 30 μm).

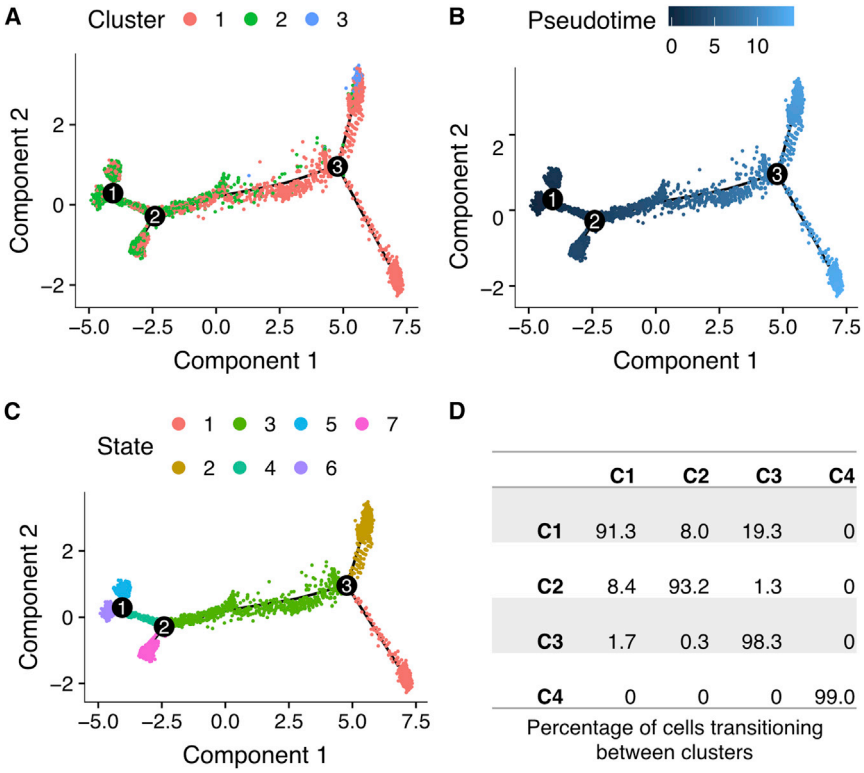
(D and E) Flow cytometry (D) was also used to quantify mean fluorescence intensity (MFI) for TMRM staining between EVP (red) and D (green), with EVP again demonstrating higher intensity of staining compared to D (n = 3; *p < 0.05; E).

(F) Violin plot of *Cyb5a* expression between EVP and D cells. The x axis shows the cluster type, and the y axis shows expression levels (normalized counts).

A redundant finding in our gene expression analysis of EVP cells was the group of genes involved in metabolism. We therefore explored cell metabolism by comparing mitochondrial activity using flow cytometry and immunofluorescence between the populations EVP and D. Using tetramethylrhodamine methyl ester (TMRM), a marker of mitochondrial membrane potential and a gold standard in measuring mitochondrial activity (Scaduto and Grotyohann, 1999), we clearly demonstrated that the EVP population had significantly higher mitochondrial activity compared to D, based on TMRM staining under immunofluorescence Z stack (Figure 4C) and geometric mean fluorescence intensity (MFI) (44,669 ± 2,354 EVP versus 29,732 ± 543 D; $p < 0.05$; n = 3; Figures 4D and 4E). On confocal microscopy in sorted cells, we could clearly observe that the increased TMRM staining was related to increased number of mitochondria (Figure 4C). In addition, within our scRNA-seq and bulk RNA-seq datasets, we observed a 2.9- and 3.0-fold increase, respectively, of *Cyb5a* (mitochondrial membrane protein) expression in EVP cells compared to D, further validating the notion of higher mitochondrial content within the EVP population (Figure 4F). These changes in mitochondrial number and activity are likely to reflect variation in metabolic activity in EVP compared to D cells and validate the scRNA gene expression. Overall, these findings validate key significant changes in gene expression between these two populations of endothelial cells, highlighting different properties in mitochondrial activity, adhesion, and interaction with extracellular matrix proteins.

Cell Cycle Phase Classification

EVP cells are more likely to be a quiescent population compared to a more actively proliferating population, such as the D cells (Patel et al., 2017), and the prediction of cell cycle phases can help to stratify the cell types. To classify the cell cycle phase of each individual cell, we used the *cyclone* method to calculate the G1, G2M, and S scores for each cell using a machine learning dataset trained on the mouse cell cycle (Lun et al., 2016; Scialdone et al., 2015). Our data revealed that the vast majority of cells (95.2%) exhibited a G1 quiescent phenotype, 4% were classified as G2M, and 0.8% were classified as S phase. We asked whether specific clusters were more quiescent or proliferative than others and observed an enrichment of proliferating cells (G2M) in clusters one (D: 5.9%; $p = 1.83 \times 10^{-26}$ [Pearson's chi-square test]) and three (D′: 19.5%; $p = 2.86 \times 10^{-68}$) compared to the EVP population (0.7%) in cluster two. Furthermore, only 0.2% of EVP cells were in S phase, indicating these cells are more quiescent than the D (1.1%; $p = 6.15 \times 10^{-5}$) or D′ (4.3%; $p = 1.52 \times 10^{-13}$)



	C1	C2	C3	C4
C1	91.3	8.0	19.3	0
C2	8.4	93.2	1.3	0
C3	1.7	0.3	98.3	0
C4	0	0	0	99.0

Figure 5. Single-Cell Trajectory Analysis Using Monocle 2 and scGPS

Pseudotime analysis of single cells using Monocle 2.

(A–C) Cells on the tree are colored by cluster assignment (A), pseudotime (B), or state (C). The top 500 genes with the highest variability in expression were used to construct the pseudotime tree. The arrangement of cells on the tree (B) shows the cells on the left side of the tree (dark blue) are less differentiated than those on the right side (light blue). Overlaying the cluster information (A) shows that these cells correspond to the progenitor cells in cluster two (EVP) and clusters one and three (D and D′), which display a more differentiated endothelial expression profile. Cell states (C) are calculated by Monocle 2 and classify the cells that are on the same branch as being most similar based on their pseudotime values.

(D) scGPS was used to calculate the proportion of cells with the potential to transition from one subpopulation to another. The transition scores for each, representing the percentage of cells with transcriptional potential to transition, reveal predominantly stable populations at each stage of differentiation.

populations, reflecting the reported quiescence as progenitors (Figures S6A and S6B).

Cell Fate Trajectory Analysis of Vascular Progenitors

We previously showed that EVP and D cells in whole tissue exist as a continuum of differentiation states. To investigate the transition of the EVP cells to the terminally differentiated D state at single-cell resolution, we performed a cell fate trajectory and pseudotime analysis using the Monocle 2 software (Qiu et al., 2017) for cells in clusters 1–3. Cluster four, which we classified as muscle cells, was excluded from this analysis, allowing us to focus on endothelial cells only. Our trajectory analysis revealed a continuum of cells with three distinct branch points, showing a root corresponding predominantly to the EVP population (cluster two) and a terminal population corresponding to D cells (cluster one; Figures 5A and 5B). States 1, 2, 3, and 7, comprised mainly of EVP cells, also contained a smaller proportion of D cells (Figures 5B and 5C; Table 2). These less differentiated D cells may represent the "transit-amplifying" (TA) population identified in our previous study (Patel et al., 2017). We observed a branch point arising mainly from cluster one (state 4; Figure 5C) that split into two branches: one comprised of D cells (state 5) and another comprised of cells from clusters one and three (state 6). To determine whether the individual states represent important subpopulations within the D and EVP populations, we tested for key gene signatures using differential expression between pairs of states. However, this did not reveal sufficiently distinct signatures to allow robust cell subtype classification beyond the root (EVP) and tip (D and D′) classifications.

Furthermore, when we overlaid the cell cycle phase information for each individual cell onto the trajectory tree, we observed a higher proportion of cells in G2M in trajectory states corresponding to more differentiated cells. This finding is consistent with our initial cell cycle analysis of each cluster. No batch effects were observed between the three pooled libraries (Figures S6C–S6E).

The high resolution of scRNA-seq enabled us to estimate the proportion of cells with the potential to transition from one cell state to another. Using scGPS (https://github.com/IMB-Computational-Genomics-Lab/scGPS) to perform an unbiased trajectory analysis, we calculated a transition score, or potential, of cells in each cluster based on gene expression signatures and cluster assignments. Our results revealed the majority of cells (>90%) in each subpopulation were stable within each cluster, indicating each population phenotype is largely fixed (Figure 5D). Notably, we detected small subpopulations of EVP (8.4%) and D (19.3%) with the potential to transition to D and D′, respectively, possibly reflecting an intermediate cell state toward differentiated endothelial cells. Our data also revealed that 8% of D cells transcriptionally lean toward the EVP state. This suggests again a small proportion of cells exist that are transcriptionally similar to both EVP and D populations and may represent the transit-amplifying cell population that we have described previously (Donovan et al., 2019).

DISCUSSION

Endothelial heterogeneity is a more and more accepted hallmark of blood vessel biology, with subpopulations identified across or

Table 2. Proportion of Cells in Each "State" as Determined by Pseudotime Analysis in Monocle 2

	Cluster			
State	1	2	3	4
1	6.2	17.3	0	0
2	2.7	3.3	0	0
3	4.0	5.6	0	0
4	6.8	2.0	0	0
5	19.1	0	0	0
6	9.3	0.5	2.7	1.3
7	4.8	14.5	0	0

within vascular beds. Despite major progress in determining the biology of these populations, an unbiased approach to delineating the endothelial heterogeneity has not yet been undertaken. To better understand this heterogeneity during homeostasis in adults, we here studied the endothelium of the largest arterial structure, the aorta, via scRNA-seq. We identified three endothelial populations, one suggestive of a progenitor population that displayed a more mesenchymal phenotype, quiescence, and high mitochondrial content, and which preceded two mature differentiated endothelial populations in pseudotime trajectories. These findings shed light on the homeostasis of the vascular endothelium.

Previous studies have reported on single-cell sequencing efforts, including endothelial cells in cancer (Sun et al., 2017) or from whole organ (e.g., heart) cluster analysis (Skelly et al., 2018). However, to date, a detailed analysis of homeostatic endothelium is lacking. We previously reported an endothelial hierarchy in the adult vasculature that is governed by an EVP that can give rise to a differentiated D population (Patel et al., 2017). This relationship between the endothelial subpopulations has been substantiated by *in vivo* lineage tracing and functional assays demonstrating that EVP cells have self-renewal and engraftment capacity. In this previous work, EVP and D cells were initially defined through observation of levels of common surface markers, such as CD31 and VEGFR2. The unbiased nature of the present study further strengthens these previous findings. We observed extremely high and significant correlation of gene expression between EVP and D and the respective clusters of endothelial cells identified here, strongly indicating that these are representations of the same populations. Sequencing depth and average read per cell reached saturation, ensuring that the observed heterogeneity reflected biological phenomena (Figure S1). Many of the genes upregulated within cluster two, that we termed the EVP, are identified as being crucial in maintaining stem cell function, phenotype, and quiescence, such as *Sox9* and *Il33* and the cyclin-dependent kinase inhibitor family (Marqués-Torrejón et al., 2013; Patel et al., 2017; Scott et al., 2010). Furthermore, metabolic activity has been demonstrated to be crucial in maintaining the self-renewal capacity of stem cell populations (Ito and Suda, 2014; Khacho and Slack, 2017; Simsek et al., 2010; Takubo et al., 2013). Within cluster two, metabolic activity was upregulated, and this population displayed a specific increase in expression of genes involved in mitochondrial activity, as validated by measuring TMRM levels as a surrogate marker of mitochondrial activity, further validating a major hierarchical difference between EVP and D cells. Our analysis also revealed in a small proportion of D cells the potential to transition in the reverse direction to EVP. However, there is little *in vivo* evidence in the context of wound healing supporting the reversion of D cells to EVP. It is more likely that these cells represent the transit-amplifying population we previously identified (Patel et al., 2017) and that the expression levels of the cell surface proteins used to identify them by flow cytometry are not strongly correlated with their mRNA expression. Fate tracing using specific genes from these transitional stages would definitively identify their role in the endothelial hierarchy.

Of note, despite the high level of correlation in gene expression between EVP and cluster two or D and cluster one, many key genes initially identified by us as determinant in these populations could not be found. Indeed, the LASSO machine learning identified a set of markers for these populations based on the single-cell data. This might be related to the preferential sequencing of highly expressed genes in scRNA-seq. It is unlikely that this is due to a lack of sequencing depth, because the dataset was sequenced to an average depth of 53,631 reads/cell and QC metrics show "sequencing saturation" and "median genes per cell" approached maximum levels (Figure S1). Similarly, other groups have also proposed the existence of resident "vascular stem cell populations." These have been identified using cell surface markers, such as CD157 in the liver (Wakabayashi et al., 2018) or activated protein C receptor in mammary tissue (Yu et al., 2016). However, to our surprise, neither of these markers could be robustly identified in our single-cell dataset and distinguish populations with the same clarity as *Il33*, *Sox9*, or *Sox18*. Certainly, further studies are warranted to identify the expression of these stem cell markers within aortic endothelial clusters.

Another key finding was the homogeneous high expression of *Pdgfrα* in EVPs. This is a well-described cell surface receptor that accurately distinguishes mesenchymal stem cell populations from both mouse and human tissue of various organs (Houlihan et al., 2012; Uezumi et al., 2014). The identification of *Pdgfrα* within cluster two in addition to a large number of extracellular matrix genes raises the potential for this population having bi-potential capability, a capacity to give rise to both endothelial and mesenchymal cells. We recently reported such bi-potential progenitor within the endothelium of the human term placenta, isolated from the vascular bed (expressing VE-cadherin) but also having high expression of mesenchymal genes, including *Pdgfrα*, and giving rise to both endothelial and mesenchymal colonies *in vitro* (Shafiee et al., 2018). An additional study during embryonic development demonstrated the crucial role for *Pdgfrα* within the endothelium in forming mesenchyme-derived structures, such as cardiac ventricles (Aghajanian et al., 2017). The resultant loss of *Pdgfrα* in the developing endothelium prevented the correct formation of ventricle structures. Therefore, this further suggests that cluster two/EVP may present as a conserved bi-potential cell in the adult endothelium that maintains this exclusive capacity and warrants further study *in vivo*.

In summary, we report an unbiased single-cell transcriptomic approach in delineating subpopulations of the homeostatic

endothelial compartment from the adult aorta. The identification of two main clusters very strongly recapitulates recently identified progenitor EVP and differentiated D populations. This provides an opportunity to understand the molecular events occurring during the transition from progenitor to differentiated cell in the resting endothelium of the adult aorta.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Tissue processing of murine aorta and FACS
  - Immunofluorescence and Microscopy
  - Metabolism Assay
  - Single-cell sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Linear modeling of transcript expression
  - Cell fate trajectory analysis
  - Gene co-expression analysis
  - Correlation analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.04.102.

ACKNOWLEDGMENTS

We would like to thank TRI Flow Cytometry facility for their assistance in the cell sorting preparation for this study and the IMB Sequencing Facility for sequencing the single-cell RNA-seq libraries. This work was supported by ARC Discovery project (DP190103187). K.K. was supported by NHMRC Career Development Fellowship (APP1125290). J.P. was supported by the ARC DECRA Research Fellowship (DE180100984). J.E.P. was supported by NHMRC Career Development Fellowship (APP1107599).

AUTHOR CONTRIBUTIONS

S.W.L. performed the scRNA-seq and computational analysis and wrote and edited the manuscript. J.P. generated cells for scRNA-seq, completed all biological validation and analysis, and wrote and edited the manuscript. S.B.A. assisted in scRNA-seq. S.-L.S. assisted in biological validation and FACS. H.Y.W. assisted in biological validation and microscopy. J.T. assisted in biological validation and metabolism assay. I.W. assisted in data analysis and editing of manuscript. J.E.P. assisted in computational analysis of scRNA-seq data and edited the manuscript. K.K. conceived and supervised the study, wrote and edited the manuscript, and generated funding.

DECLARATION OF INTERESTS

J.P. and K.K. are co-inventors relating to the isolation of endothelial progenitors from the placenta.

Received: August 22, 2018
Revised: March 7, 2019
Accepted: April 22, 2019
Published: May 28, 2019

REFERENCES

Adams, R.H., and Alitalo, K. (2007). Molecular regulation of angiogenesis and lymphangiogenesis. Nat. Rev. Mol. Cell Biol. *8*, 464–478.

Aghajanian, H., Cho, Y.K., Rizer, N.W., Wang, Q., Li, L., Degenhardt, K., and Jain, R. (2017). Pdgfrα functions in endothelial-derived cells to regulate neural crest cells and the development of the great arteries. Dis. Model. Mech. *10*, 1101–1108.

Aird, W.C. (2007a). Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms. Circ. Res. *100*, 158–173.

Aird, W.C. (2007b). Phenotypic heterogeneity of the endothelium: II. Representative vascular beds. Circ. Res. *100*, 174–190.

Aird, W.C. (2012). Endothelial cell heterogeneity. Cold Spring Harb. Perspect. Med. *2*, a006429.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B Met. *57*, 289–300.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Donovan, P., Patel, J., Dight, J., Wong, H.Y., Sim, S.-L., Murigneux, V., Francois, M., and Khosrotehrani, K. (2019). Endovascular progenitors infiltrate melanomas and differentiate towards a variety of vascular beds promoting tumor metastasis. Nat. Commun. *10*, 18.

Eichmann, A., Yuan, L., Moyon, D., Lenoble, F., Pardanaud, L., and Breant, C. (2005). Vascular development: from precursor cells to branched arterial and venous networks. Int. J. Dev. Biol. *49*, 259–267.

Emara, M., Turner, A.R., and Allalunis-Turner, J. (2014). Hypoxia differentially upregulates the expression of embryonic, fetal and adult hemoglobin in human glioblastoma cells. Int. J. Oncol. *44*, 950–958.

Fang, S., Wei, J., Pentinmikko, N., Leinonen, H., and Salven, P. (2012). Generation of functional blood vessels from a single c-kit+ adult vascular endothelial stem cell. PLoS Biol. *10*, e1001407.

Herbert, S.P., and Stainier, D.Y. (2011). Molecular control of endothelial cell behaviour during blood vessel morphogenesis. Nat. Rev. Mol. Cell Biol. *12*, 551–564.

Houlihan, D.D., Mabuchi, Y., Morikawa, S., Niibe, K., Araki, D., Suzuki, S., Okano, H., and Matsuzaki, Y. (2012). Isolation of mouse mesenchymal stem cells on the basis of expression of Sca-1 and PDGFR-α. Nat. Protoc. *7*, 2103–2111.

Iruela-Arispe, M.L., and Davis, G.E. (2009). Cellular and molecular mechanisms of vascular lumen formation. Dev. Cell *16*, 222–231.

Ito, K., and Suda, T. (2014). Metabolic requirements for the maintenance of self-renewing stem cells. Nat. Rev. Mol. Cell Biol. *15*, 243–256.

Khacho, M., and Slack, R.S. (2017). Mitochondrial activity in the regulation of stem cell self-renewal and differentiation. Curr. Opin. Cell Biol. *49*, 1–8.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122.

Marqués-Torrejón, M.A., Porlan, E., Banito, A., Gómez-Ibarlucea, E., Lopez-Contreras, A.J., Fernández-Capetillo, O., Vidal, A., Gil, J., Torres, J., and Fariñas, I. (2013). Cyclin-dependent kinase inhibitor p21 controls adult neural stem cell expansion by regulating Sox2 gene expression. Cell Stem Cell *12*, 88–100.

Newton, D.A., Rao, K.M., Dluhy, R.A., and Baatz, J.E. (2006). Hemoglobin is expressed by alveolar epithelial cells. J. Biol. Chem. *281*, 5668–5676.

Nguyen, Q.H., Lukowski, S.W., Chiu, H.S., Senabouth, A., Bruxner, T.J.C., Christ, A.N., Palpant, N.J., and Powell, J.E. (2018). Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. Genome Res. *28*, 1053–1066.

Patel, J., Seppanen, E.J., Rodero, M.P., Wong, H.Y., Donovan, P., Neufeld, Z., Fisk, N.M., Francois, M., and Khosrotehrani, K. (2017). Functional definition of progenitors versus mature endothelial cells reveals key SoxF-dependent differentiation process. Circulation *135*, 786–805.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Saha, D., Patgaonkar, M., Shroff, A., Ayyar, K., Bashir, T., and Reddy, K.V. (2014). Hemoglobin expression in nonerythroid cells: novel or ubiquitous? Int. J. Inflamm. *2014*, 803237.

Saha, D., Koli, S., Patgaonkar, M., and Reddy, K.V. (2017). Expression of hemoglobin-α and β subunits in human vaginal epithelial cells and their functional significance. PLoS ONE *12*, e0171084.

Scaduto, R.C., Jr., and Grotyohann, L.W. (1999). Measurement of mitochondrial membrane potential using fluorescent rhodamine derivatives. Biophys. J. *76*, 469–477.

Scialdone, A., Natarajan, K.N., Saraiva, L.R., Proserpio, V., Teichmann, S.A., Stegle, O., Marioni, J.C., and Buettner, F. (2015). Computational assignment of cell-cycle stage from single-cell transcriptome data. Methods *85*, 54–61.

Scott, C.E., Wynn, S.L., Sesay, A., Cruz, C., Cheung, M., Gomez Gaviro, M.V., Booth, S., Gao, B., Cheah, K.S., Lovell-Badge, R., and Briscoe, J. (2010). SOX9 induces and maintains neural stem cells. Nat. Neurosci. *13*, 1181–1189.

Shafiee, A., Patel, J., Hutmacher, D.W., Fisk, N.M., and Khosrotehrani, K. (2018). Meso-endothelial bipotent progenitors from human placenta display distinct molecular and cellular identity. Stem Cell Reports *10*, 890–904.

Simsek, T., Kocabas, F., Zheng, J., Deberardinis, R.J., Mahmoud, A.I., Olson, E.N., Schneider, J.W., Zhang, C.C., and Sadek, H.A. (2010). The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. Cell Stem Cell *7*, 380–390.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R. (2018). Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. Cell Rep. *22*, 600–610.

Sun, Z., Wang, C.Y., Lawson, D.A., Kwek, S., Velozo, H.G., Owyong, M., Lai, M.D., Fong, L., Wilson, M., Su, H., et al. (2017). Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells. Oncotarget *9*, 10945–10961.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Takubo, K., Nagamatsu, G., Kobayashi, C.I., Nakamura-Ishizu, A., Kobayashi, H., Ikeda, E., Goda, N., Rahimi, Y., Johnson, R.S., Soga, T., et al. (2013). Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. Cell Stem Cell *12*, 49–61.

Uezumi, A., Fukada, S., Yamamoto, N., Ikemoto-Uezumi, M., Nakatani, M., Morita, M., Yamaguchi, A., Yamada, H., Nishino, I., Hamada, Y., and Tsuchida, K. (2014). Identification and characterization of PDGFRα+ mesenchymal progenitors in human skeletal muscle. Cell Death Dis. *5*, e1186.

Wakabayashi, T., Naito, H., Suehiro, J.I., Lin, Y., Kawaji, H., Iba, T., Kouno, T., Ishikawa-Kato, S., Furuno, M., Takara, K., et al. (2018). CD157 marks tissue-resident endothelial stem cells with homeostatic and regenerative properties. Cell Stem Cell *22*, 384–397.e6.

Yu, Q.C., Song, W., Wang, D., and Zeng, Y.A. (2016). Identification of blood vascular endothelial stem cells by the expression of protein C receptor. Cell Res. *26*, 1079–1098.

Zhao, Q., Eichten, A., Parveen, A., Adler, C., Huang, Y., Wang, W., Ding, Y., Adler, A., Nevins, T., Ni, M., et al. (2018). Single-cell transcriptome analyses reveal endothelial cell heterogeneity in tumors and changes following antiangiogenic treatment. Cancer Res. *78*, 2370–2382.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PE Rat Anti-Mouse VEGFR2	Becton Dickinson	555308; RRID:AB_395271
PE-Cy7 Rat Anti-Mouse CD31	Becton Dickinson	561410; RRID:AB_10612003
Alexa Fluor 647 Rat Anti-Mouse CD34	Becton Dickinson	560230; RRID:AB_1645200
BV450 Rat Anti-Mouse Lineage Cocktail	BioLegend	133310; RRID:AB_10768377
Alexa Fluor 488 Rat Anti-Mouse CD144	eBioscience	5016730; RRID:AB_10597442
Goat Polyclonal Anti-Decorin	R&D	AF1060; RRID:AB_2090386
Rabbit Polyclonal Anti-Periostin	Abcam	ab14041; RRID:AB_2299859
Donkey Anti-Goat Alexa Fluor 488	Invitrogen	ab150129; RRID:AB_2687506
Donkey Anti-Rabbit Alexa Fluor 488	Invitrogen	ab150073; RRID:AB_2636877
Tetramethylrhodamine, Methyl Ester	Life Technologies	T668
Chemicals, Peptides, and Recombinant Proteins		
Collagenase Type I	GIBCO	17100017
Dispase II	GIBCO	17105041
Deoxyribonuclease I	Sigma-Aldrich	D5025-375KU
Critical Commercial Assays		
Chromium Single Cell 3′ Gel Bead and Library Kit v2	10X Genomics	PN-120237
KAPA Library Quantification Kit	Roche	KR0405 – v8.17
Deposited Data		
scRNA-seq data	This paper	E-MTAB-7149
Bulk RNA-seq data	Donovan et al., 2019	E-MTAB-7148
Experimental Models: Organisms/Strains		
Mouse: C57BL/6JArc	Animal Resources Centre Perth	JAX: 000664
Software and Algorithms		
Cellranger	10x Genomics	2.1.0
Ascend	https://github.com/IMB-Computational-Genomics-Lab/ascend	0.5
Scran	Lun et al., 2016	1.8.4
DESeq	Anders and Huber, 2010	1.32.0
Monocle2	Qiu et al., 2017	2.8.0
scGPS	https://github.com/IMB-Computational-Genomics-Lab/scGPS	0.9.9
WGCNA	Langfelder and Horvath, 2008	1.63
FV31S-SW	Olympus	Ver2.3.1
ArrayExpress	https://www.ebi.ac.uk/arrayexpress	Accession numbers E-MTAB-7149 and E-MTAB-7148

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Professor Kiarash Khosrotehrani (k.khosrotehrani@uq.edu.au).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All mice were treated in accordance with University of Queensland ethics approvals and guidelines for care of experimental animals. C57BL/6 mice (WT) females were obtained from the Animal Resources Centre (Perth, Western Australia).

METHOD DETAILS

Tissue processing of murine aorta and FACS

Aortas were dissected and digested for 30 mins at 37°C in 1mg/ml collagenase I (GIBCO, Life Technologies, NY, USA), 1mg/ml Dispase (GIBCO, Life Technologies, NY, USA), 150 μg/ml DNase-I (Sigma-Aldrich, St Louis, MO, USA) before passing the suspension through a 70 μM cell strainer. Dissociated single cells in PBS/BSA/EDTA were then incubated with various antibody combinations for multi-parameter flow acquisition and analysis. The following combinations of antibodies were used to assess the endothelial hierarchy populations: Rat anti-mouse VEGFR2 PE, CD31 PE-Cy7 and CD34 Alexa647 (Becton Dickinson, NJ, USA), Rat anti-mouse Lineage cocktail BV450 (Biolegend), Rat anti-mouse CD144 FITC (eBioscience). Fluorescence-activated cell sorting (FACS) was performed on a FACSAria cell sorter (Becton Dickinson, Franklin Lakes, NJ, USA). Doublets were gated out prior to isolation. Isolated cells were then cytospun onto Superfrost Plus Microscopy slides (Thermo Scientific, MA, USA) for immunofluorescence analysis.

Immunofluorescence and Microscopy

For this study, primary antibodies used included goat anti-decorin (1:50) and rabbit anti-periostin (1:100) (all from Abcam, MA, USA). Excess and unbound antibody was then removed with 3x 5mins washes in a solution containing 1x PBS/0.1% Tween-20 (Amresco, Solon, Ohio, USA). Secondary antibody conjugated with Alexa Fluor 488 (Invitrogen, Carlsbad, CA, USA) were used for fluorescence detection. Briefly, slides were incubated with secondary antibodies for 40mins at room temperature. Excess antibody was removed by 3x washes in PBS/0.1% Tween-20. Nuclear staining was revealed in specimens mounted with ProLong® Gold mounting media containing DAPI (Invitrogen, Carlsbad, CA, USA). Confocal images were acquired with a Zeiss LSM 710 microscope Images were obtained at 63x.

Metabolism Assay

Filtered aortic cell suspensions were incubated with 20 nM tetramethylrhodamine, methyl ester (TMRM, Thermofisher) for 20 min at 37°C in the dark with agitation. Cells were then washed in ice-cold PBS/BSA/EDTA. Cells were then stained with antibodies against cell surface markers for 30 min in the dark on ice. Excess antibodies were washed off with PBS/BSA/EDTA buffer. The fluorescence of mitochondrial bound TMRM was determined by flow cytometry and immunofluorescence, with excitation at 561 nm and emission at 585 ± 21 nm. Cell populations were analyzed using CytoFLEX S analyzer (Beckman Coulter) followed by further analyses using FlowJo v7-10 software (TreeStar).

Single-cell sequencing

FACS-sorted single-cell suspensions were used to generate barcoded single-cell 3′ cDNA libraries for each sample pool with the Chromium Single Cell 3′ Gel Bead and Library Kit v2 (10x Genomics). Libraries were assessed with an Agilent BioAnalyzer High Sensitivity DNA chip, pooled, and quantified with qPCR (KAPA Library Quantification Kit). Denatured libraries were loaded onto an Illumina NextSeq-500 and sequenced using a 150-cycle High-Output Kit to an average depth of 53,631 reads/cell. To process the sequencing data, we used the 10x Genomics *cellranger* pipeline (v2.1.0), comprising the *mkfastq*, *count* and *aggr* stages. Using *cellranger mkfastq*, raw base call files were demultiplexed into sample-specific FASTQ files. Sample-specific FASTQ files were quality controlled and aligned to the mouse reference genome (mm10) using STAR aligner (Dobin et al., 2013) during the *cellranger count* stage. Aligned reads were filtered for valid cell barcodes and unique molecular identifiers. An aggregated, between-sample normalized gene expression matrix was generated using *cellranger aggr*.

The aggregated single-cell gene expression data generated by *cellranger* was used as the input for the *ascend* analysis pipeline (v0.5; https://github.com/IMB-Computational-Genomics-Lab/ascend). Expression levels for each transcript were determined using the number of unique molecular identifiers (UMI) assigned to the transcript. Quality control and filtering steps were performed to remove outlier cells and genes. Cells were excluded if the library size, or number of expressed genes exceeded 3 median absolute deviations, or if mitochondrial or ribosomal reads accounted for than 20% or 50% of the reads respectively. The relative log expression (RLE) method was used to normalize cell-cell expression variation, and ribosomal and mitochondrial control genes were subsequently excluded from the analysis.

Principal component analysis was performed on the filtered and normalized gene expression matrix, and the first 20 PCs that explained the majority of variance in the data were retained. Unsupervised hierarchical clustering of individual cells was applied using the method previously described (Nguyen et al., 2018). To visualize the patterns of gene expression in each cell in each cluster, t-distributed stochastic neighbor (t-SNE) projections were generated using the PCA-reduced data.

Differential expression analysis was performed with DESeq (Anders and Huber, 2010) in *ascend* between (i) each cluster compared to the remaining cells, or (ii) between pairs of specific clusters. Differences in gene expression were considered significant if the adjusted P value was below the multiple-testing threshold of 0.01 (Benjamini and Hochberg, 1995) and the absolute log2 expression fold change was ≥ 2.

Cell cycle phase classification of each cell (n = 6,111) was performed using the *cyclone* function implemented in the scran package (Lun et al., 2016), with a dataset trained on mouse cell cycle gene expression.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise noted, all data are represented as mean ± standard error of mean (SEM). Indicated sample sizes (n) represent biological replicates including individual samples, these have been listed in the results section and figure legends where appropriate. Statistical significance where appropriate was determined in GraphPad Prism 7 software by using the Mann-Whitney t-test (non-parametric). For validation of sequencing data, a significance threshold of $p < 0.05$ was used.

Linear modeling of transcript expression

Linear regression was used to test the expression of each individual gene (n = 17,323) against the cluster classification of the 5,867 cells comprising clusters 1 and 2 only. The overall model P value was adjusted for multiple testing using a Bonferroni threshold calculated as $0.05/17{,}323 = 2.89 \times 10^{-6}$ and transcripts with an adjusted P value below the threshold were considered statistically significant.

Cell fate trajectory analysis

Monocle2 (Qiu et al., 2017) was used to order cells in pseudo-time based on the top 500 genes with the highest variation in expression across the 6,031 cells comprising clusters 1-3. The most variable genes were determined using a $\log_2$(counts + 1)-transformed dataset. The LASSO regression procedure in scGPS (https://github.com/IMB-Computational-Genomics-Lab/scGPS) was used to calculate the deviance coefficient for each of the genes defining each cluster. For the scGPS input, we used the significant differentially expressed genes that were obtained by comparing cluster *x* to the remaining clusters, a vector of cluster information for each cell, and $\log_2$(counts + 1)-transformed expression data.

Gene co-expression analysis

To establish gene co-expression networks underlying the EVP and D populations in our single-cell data, we applied weighted gene co-expression network analysis as implemented in the WGCNA package for R (Langfelder and Horvath, 2008).

Correlation analysis

Mean expression values for scRNA-seq clusters and bulk RNA-seq samples was used to perform pairwise Spearman's correlations. Correlations were performed with known gene markers for EVP and D cells derived from previous studies on normal (Patel et al., 2017) and tumor-derived populations (Donovan et al., 2019), and these were merged with differentially expressed genes defining the scRNA-seq clusters. This yielded a common set of 2,105 genes between all samples that were taken forward for pairwise correlation analysis.

DATA AND SOFTWARE AVAILABILITY

Single-cell RNA sequencing data was deposited in ArrayExpress (https://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7149 and bulk RNA sequencing data under accession number E-MTAB-7148.